The Institute of Paper Chemistry

Appleton, Wisconsin

Doctor's Dissertation

An Investigation of the Vibrational Spectra of the Cellodextrins

Kenneth Paul Carlson

January, 1979

AN INVESTIGATION OF THE VIBRATIONAL SPECTRA OF THE CELLODEXTRINS

A thesis submitted by

Kenneth Paul Carlson

B.A. (Chem. and Phys.) 1973, Coe College - Cedar Rapids, Iowa

M.S. 1975, Lawrence University

in partial fulfillment of the requirements
of The Institute of Paper Chemistry
for the degree of Doctor of Philosophy
from Lawrence University
Appleton, Wisconsin

January, 1979

TABLE OF CONTENTS

SUMMARY

The purpose of this thesis is to investigate the vibrational spectra of the cellodextrins with particular emphasis on the sensitivity of the vibrational spectra to conformational changes about the glycosidic linkage(s). A better understanding of the vibrational spectra of the cellodextrins will aid the interpretation of the vibrational spectra of the various polymorphs of cellulose. It will also hopefully lead to a better understanding of the nature of the polymorphy.

The vibrational spectra of several disaccharides were investigated by means of normal coordinate analysis using a set of force constants previously developed for α and β glucose. The agreement between the calculated and observed frequencies was very good for cellobiose, methyl β-cellobioside, xylobiose, α-lactose and maltose. Normal coordinate calculations were also performed for various conformations of cellobiose to determine the effect of conformational changes on the calculated frequencies. The calculations indicate that the differences between the observed vibrational spectra of cellulose I and cellulose II can be accounted for by differences in the torsional angles, ϕ and ψ, about the glycosidic linkages. Conformational differences at the hydroxymethylene positions resulted in changes in the calculated frequencies not consistent with the observed differences between the observed frequencies for cellulose I and cellulose II.

The vibrational spectra of the cellodextrins were investigated using normal coordinate calculations for cellotetraose structures. The differences between the observed spectra of cellulose I and cellulose II can also be accounted for by differences in the torsional angles, ϕ and ψ, at the glycosidic linkages based on the cellotetraose calculations.

The O-H stretching region of the vibrational spectra of the disaccharides and the cellodextrins has been interpreted in terms of intermolecular and intramolecular hydrogen bonding. The results of the interpretation are consistent with the conclusions of the investigations based on normal coordinate calculations.

INTRODUCTION

The forest is the world's major renewable resource. As with all of nature, trees exhibit various levels of structure. The tree, upon first inspection is composed of a trunk, limbs, twigs, and leaves. The trunk and limbs have a layer of bark, a cambium, and the core. This core is composed of many fibers which also have a well defined structure. There are three distinct layers of fibrils within these fibers. The fibrils are constructed of cellulose chains having a definite arrangement. The cellulose chains are composed of anhydroglucose units linked together by β-1$\rightarrow$4 glycosidic linkages. It is the spatial arrangement and motions of these anhydroglucose units which is of interest in this thesis work.

Cellulose has the property of existing in various forms exhibiting different physical properties. Cellulose I, the native form, can be transformed into cellulose II, or regenerated cellulose, either by mercerization or by dissolution followed by precipitation. One of the differences among the forms of cellulose lies in their physical reaction with light. These differences result from differences in the natural modes of vibration for these forms and can be observed in their vibrational spectra. The purpose of this thesis work is to aid the interpretation of the differences in the vibrational spectra.

Recent advances in digital computers and laser Raman spectrometers have allowed significant strides in the interpretation of the vibrational spectra of several carbohydrates. It is now possible to investigate the vibrational spectra of several disaccharides using normal coordinate analysis. The information gained in these calculations may then form a basis for the interpretation of the vibrational spectra of the cellodextrins. This information, coupled with normal coordinate calculations for cellotetraose, will aid the interpretation of the differences in the vibrational spectra of cellulose I and cellulose II.

SECTION I

BACKGROUND

EARLY INVESTIGATIONS

Interest in the Raman and infrared spectra of the carbohydrates has resulted in the compilation of a large amount of data. Early investigations were limited to the group frequency approach due to the large volume of computations inherent in a complete theoretical treatment. The group frequency approach assumes that certain molecular moieties vibrate at characteristic frequencies independently from the rest of the molecule. The first step in developing the group frequency approach is the investigation of the vibrational spectra of simple compounds having atomic groupings similar to the carbohydrates. These preliminary investigations then form a basis upon which to interpret the more complex vibrational spectra. This methodology is still used today as a useful analytical technique. The region of the vibrational spectra of the carbohydrates which yields the most information concerning molecular configuration lies below 1200 cm^{-1} and is referred to as the "fingerprint region." Unfortunately the group frequency approach is severely limited when applied to this region. This is due to the high degree of coupling between the various types of vibration.

The application of the group frequency approach was utilized for a number of years as the method of interpreting the vibrational spectra of the carbohydrates. Kuhn[1] published the vibrational spectra of seventy-nine carbohydrates with an interpretation of only a few characteristic frequencies. Differences in the infrared spectra between 730 and 960 cm^{-1} for a number of pyranose sugars were interpreted as differences between the α and β anomers by Barker.[2,3] The same region was also interpreted for several deoxy-compounds by Barker.[4] The literature on the interpretation of the vibrational spectra of the carbohydrates prior

to 1957 has been reviewed by Neely.[5] More recent reviews have been published by Spedding[6] and Tipson.[7]

The results of these many investigations indicate the limitation of the group frequency approach. The group frequency approach is useful in that it allows one to make general comments as to the chemical composition of the compound based on certain characteristic frequencies. However, the majority of vibrational modes are composed of numerous atomic motions within a molecule. At this point the assumption of molecular moieties vibrating independently from the rest of the molecule is no longer valid. This has been shown to be the case for the vibrations having frequencies below 1200 cm^{-1} in the carbohydrates.

APPLICATIONS OF NORMAL COORDINATE ANALYSIS

Normal coordinate analysis (NCA) is available for the calculation of vibrational frequencies and the associated molecular motions for any given molecule. The application of NCA to even relatively simple molecules, however, has been limited by the volume of computational work which is required. The advent of high speed digital computers allowed the application of NCA to these simple molecules and, in time, to more complex carbohydrates. In 1963, Schachtschneider and Snyder[8,9] published a definitive work on the application of high speed digital computers to the NCA of a homologous series of compounds. They were able to fit the calculated frequencies to the observed frequencies for a series of n-paraffins by adjusting a set of force constants common to the group of molecules. Snyder and Schachtschneider[10] extended their work to other saturated hydrocarbons including cyclohexane. This work demonstrated the applicability of NCA to a series of large complex molecules. It also established a procedure for obtaining, through refinement, a set of force constants for the series of compounds.

The NCA of several ether compounds, including tetrahydropyran, was reported in 1967 by Snyder and Zerbi.[11] A NCA of cyclohexane tetrahydropyran and other related oxanes has been reported by Pickett and Strauss[12] utilizing the method of Schachtschneider and Snyder. These studies also demonstrated the applicability of NCA in the interpretation of the vibrational spectra of relatively complex molecules.

Workers at Case Western Reserve have published a series of articles using NCA in the interpretation of the vibrational spectra of several carbohydrates. The NCA of α-D-glucose was reported by Vasko[13] in 1972. This NCA was performed utilizing force constants from various literature sources with no attempt at refinement. This work produced a more detailed interpretation of the vibrational spectrum of α-D-glucose than previously reported. However, the use of literature values for force constants without any refinement makes the results somewhat questionable. There is also some question concerning the assignment of observed fundamental frequencies. Further work on D-glucose has been reported by Cael.[14] Particular attention was paid to vibrations sensitive to anomeric configuration. This work was extended to some polysaccharides, including a cellulose I model, by Cael.[15,16] Again, force constants were taken from the literature without refinement. In order to facilitate the calculations for the cellulose I model, certain assumptions were made concerning its geometrical structure. The major assumption was that the cellulose I model possessed an axis of twofold screen symmetry. The representation of cellulose I with a model having twofold screw symmetry has been questioned on the basis of x-ray diffraction studies.[17] Cael reached the general conclusion that above 1200 cm^{-1} the vibrations are localized within an anhydroglucose unit but that significant coupling exists between neighboring residues for frequencies below 1200 cm^{-1}.

Recently, Zhbankov[17] has published an NCA of cellobiose. It appears that he used the same set of force constants used by the workers at Case Western Reserve[13-16] and the crystal structure for cellobiose published by Chu and Jeffrey[18] as the basis for his calculations. A software package similar to that published by Schachtschneider and Snyder[8,9] was developed without the option for refinement of the force constants. The results were not presented in detail; the normal modes of vibration were discussed in general terms. This work demonstrates that NCA can be a useful tool in the investigation of the vibrational spectra of disaccharides and it presents a more detailed interpretation of the vibrational spectrum of cellobiose than was previously available. However, it is difficult to assess the results of this work due to the general manner of their presentation.

NORMAL COORDINATE ANALYSIS OF MODEL COMPOUNDS

The vibrational spectra of several model compounds have been interpreted in thesis work at The Institute of Paper Chemistry. These compounds are the 1,5 anhydropentitols, the acyclic alditols, the pentoses, the glucopyranoses, and the inositols. The purpose of these studies was to establish a basis for the interpretation of the vibrational spectra of the carbohydrates in general and cellulose in particular.

Pitzner[19,20] investigated the vibrational spectra of the 1,5 anhydropentitols using NCA. This class of compounds was chosen because it represents the first step from tetrahydropyran to the saccharide structures. Pitzner interpreted and assigned the solid state spectra for 1,5 anhydroxylitol, 1,5 anhydroribitol, and 1,5 anhydroarabinitol. A nonlinear least squares technique based on the method of Fletcher and Powell[21] was employed for the refinement of force constants

after the linear Gauss-Newton-Rapson method used by Schachtschneider and Snyder[8,9] proved unsuccessful. Good agreement was obtained between the calculated and observed frequencies. Spectral differences were attributed to conformational or "G matrix" effects. Calculated frequencies were also obtained for the compounds studied in which the rings were "flipped," (i.e., the conformation of the rings was changed from the C1 to the higher 1C conformation). The calculated frequencies for these alternative conformations agreed with "new" bands observed in the solution spectra of the compounds. This agreement was taken as support for the set of force constants obtained.

This work was followed by the investigation of several acyclic alditols by Watson.[22] A set of force constants was developed for erythritol, ribitol, and xylitol which gave good agreement between the observed and calculated frequencies. This force field was then used to predict the spectrum of another pentitol, namely D-arabinitol, which had not been included in the refinements. The agreement between the observed and calculated frequencies for this compound was taken as evidence in support of the developed set of force constants. This agreement also demonstrated the predictive capability of a carefully developed set of force constants.

The next class of compounds to be studied was the pentose sugars investigated by Edwards.[23] This set of compounds was a progression from the 1,5 anhydropentitols in that they include an anomeric hydroxyl. Edwards found that the inclusion of this anomeric hydroxyl substantially complicated the refinement of a set of force constants which would provide good agreement between the observed and calculated frequencies. He was able, however, to obtain a set of force constants which would yield good agreement between the calculated and observed frequencies for the pentose sugars β-arabinose, β-lyxose, and α-xylose.

By using this set of force constants to predict the vibrational spectra of α and β methyl xylopyranoside, Edwards was able to demonstrate the predictive capability of force constants for similar compounds.

The next step in this progression has recently been completed by Wells[24] in his investigation of the vibrational spectra of the hexoses. A forty-six parameter force field was developed which allowed good agreement between the calculated and observed frequencies for α-D-glucose, β-D-glucose, and several deuterated analogs. This set of force constants was also applied to α-mannose, β-mannose, α-galactose, and β-galactose with good agreement between the calculated and observed frequencies. Wells also investigated the effect of hydroxymethylene conformation on the calculated frequencies for glucose, galactose, and mannose. The comparison of the calculated frequencies for the three hydroxymethylene conformations with the observed frequencies allowed the prediction of a preferred conformation which was in agreement with other observations for glucose and galactose. The comparison for mannose yielded uncertain results.

Another investigation has recently been completed by Williams[25] in which the vibrational spectra of the inositols were studied. The symmetry of these compounds allowed a much more precise assignment of the calculated frequencies to the observed frequencies. Since fewer force constants were required, due to the lack of a ring oxygen, Williams was able to assess the extent to which second order effects perturb the experimentally observed frequencies of the fundamental modes of vibration. This also allowed him to make a critical evaluation of the assumptions inherent in the normal coordinate calculations. He was able to establish that second order effects were the major source of error between the calculated and observed frequencies and that the assumptions inherent in the normal coordinate calculations are valid.

The series of investigations of the vibrational spectra of the model compounds has formed the foundation for the interpretation of the vibrational spectra of these carbohydrates. One of the most ubiquitous carbohydrates is cellulose. It is a very important chemical; yet, its precise structure is the subject of much debate. One aspect of the debate centers on the fact that cellulose can exist in numerous polymorphic forms. One can identify the polymorphic form by either the x-ray diffractogram or the vibrational spectrum, either Raman or infrared, of the sample.

It is the purpose of this thesis to investigate the vibrational spectra of several compounds closely related to cellulose entering on the effect of conformation on the calculated and observed frequencies of these compounds. This will yield a better understanding and interpretation of the differences between the vibrational spectra of the various polymorphic forms of cellulose, particularly cellulose I and cellulose II. The investigation will therefore focus primarily on the vibrational regions where differences have been observed between the vibrational spectra of cellulose I and cellulose II. The vibrational spectra of several related disaccharides will be investigated initially. A better understanding of the types of motions and the interactions of these complex motions will be gained from this work. This information will be used as a guide in an investigation of the effect of conformational changes on the calculated frequencies of cellobiose. The vibrational spectra of cellotetraose will be investigated utilizing normal coordinate calculations. The results of these calculations will be applied to the interpretation of the vibrational spectra of cellulose I and cellulose II.

Other regions of the vibrational spectra, particularly the O-H stretching region, will be investigated and interpreted with regard to compound structure.

SECTION II

AN INVESTIGATION OF THE VIBRATIONAL SPECTRA OF SEVERAL DISACCHARIDES

INTRODUCTION

As discussed earlier, the vibrational spectra of several model compounds have been interpreted in thesis work at The Institute of Paper Chemistry. The purpose of these studies was to establish a basis for the interpretation of the vibrational spectra for the carbohydrates in general and cellulose in particular. This thesis extends the work to several related disaccharides which may serve as model compounds for cellulose. These include cellobiose, methyl β-cellobioside, and α-lactose. Also included in this investigation are the disaccharides maltose and xylobiose.

EXPERIMENTAL

Cellobiose, α-lactose, and maltose were obtained from a commercial source. Xylobiose and methyl β-cellobioside were obtained from sources at The Institute of Paper Chemistry.[26,27]

The observed frequencies for this investigation were the infrared and Raman spectra for the compounds being considered. The infrared spectra were recorded on a Perkin Elmer Model 621 Infrared Spectrophotometer. The Raman spectra were recorded on a Jobin Yvon Raman Spectrophotometer interfaced with a Tracor Northern TN-1500 minicomputer for data acquisition and analysis. The observed vibrational spectra for the compounds studied are illustrated in Fig. 1 through 10.

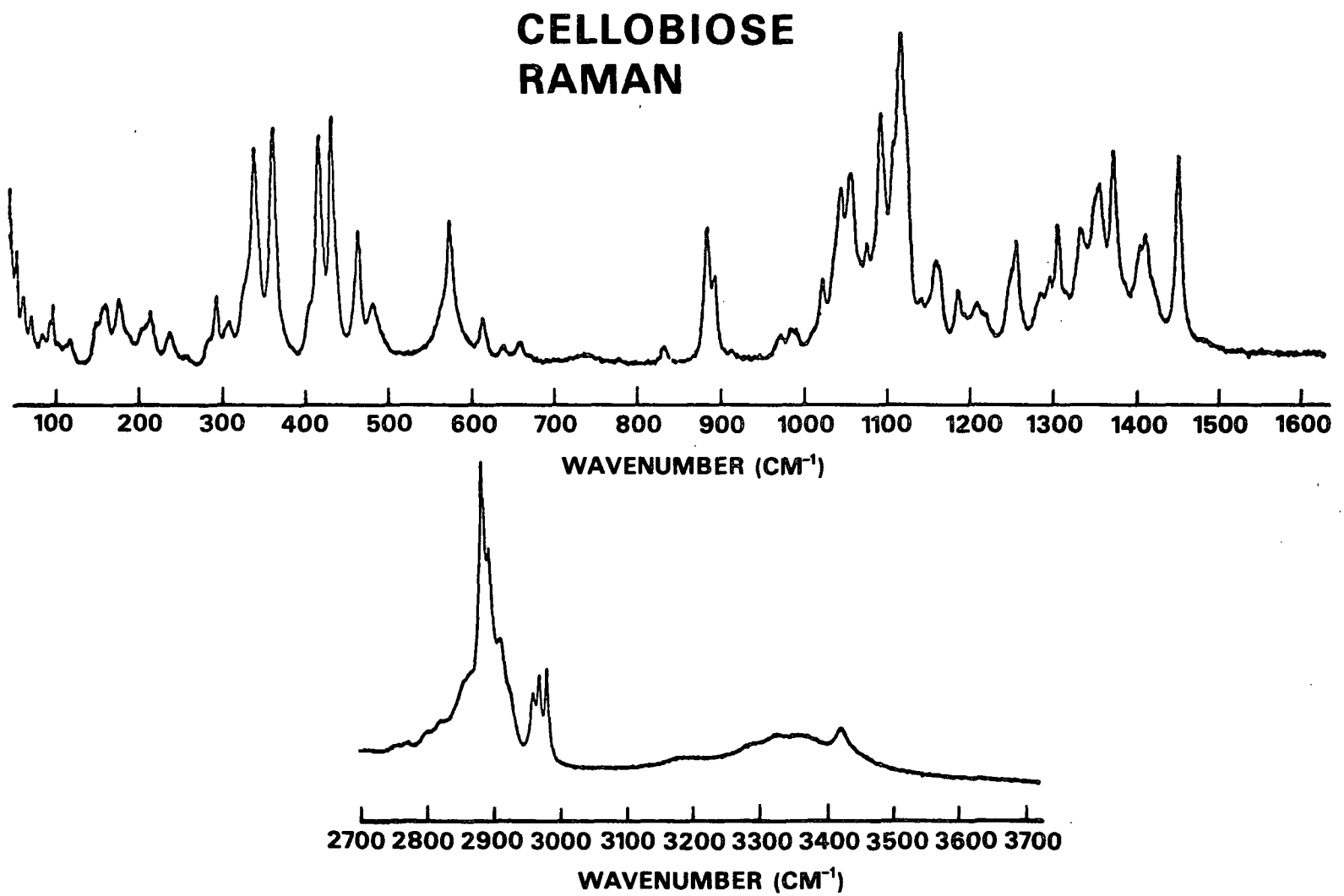


Figure 1. Raman spectrum of cellobiose.

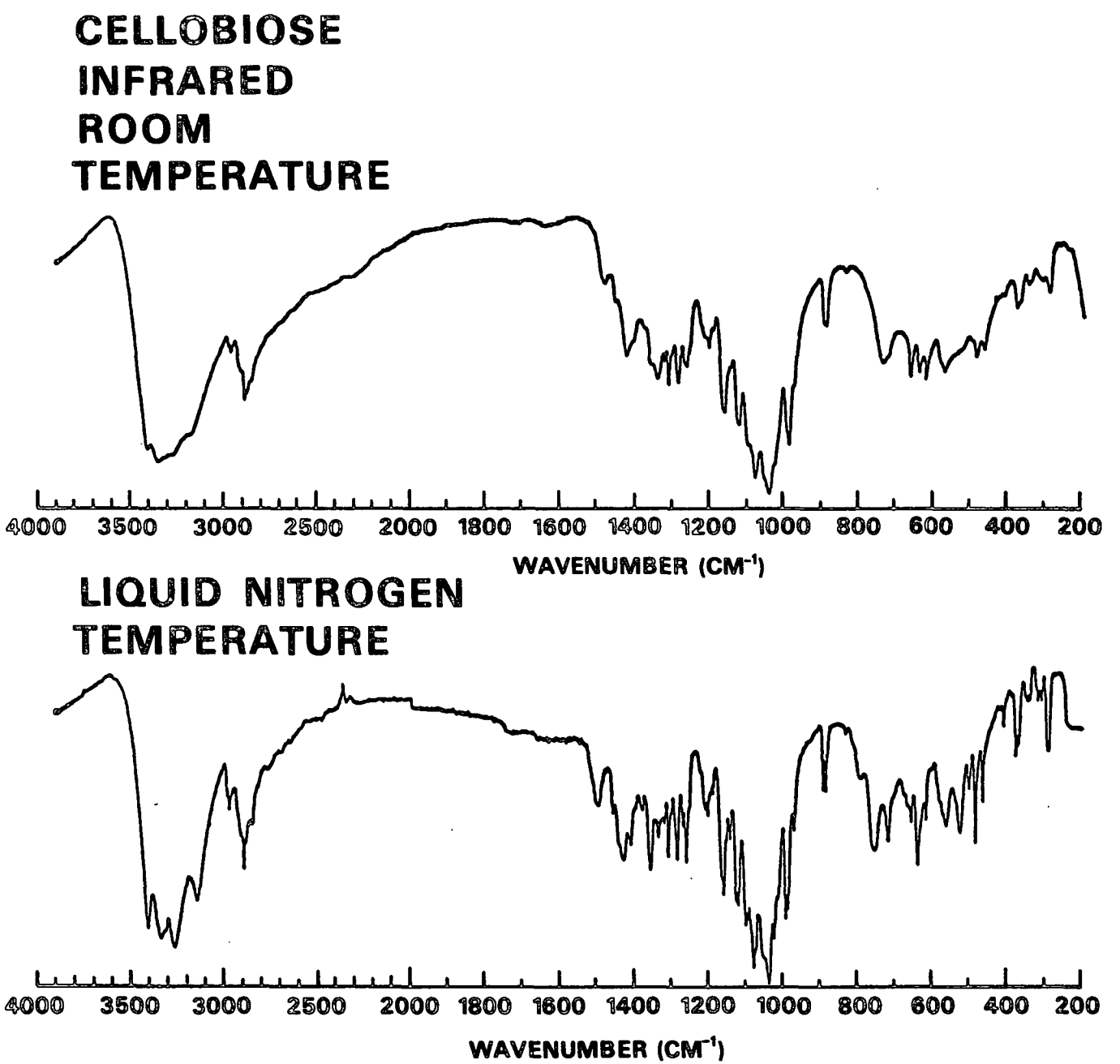


Figure 2. Infrared spectrum of cellobiose.

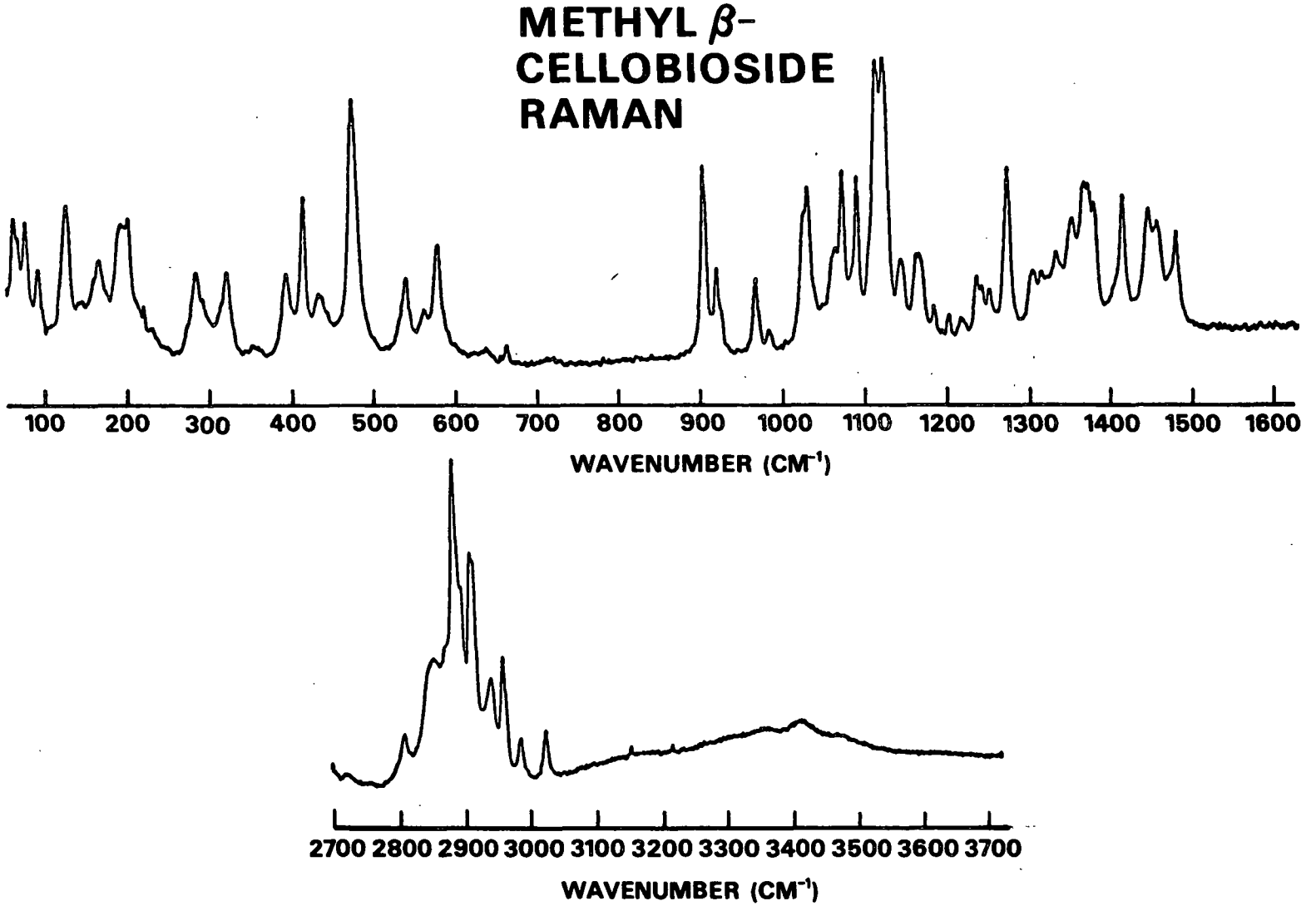


Figure 3. Raman spectrum of methyl β-cellobioside.

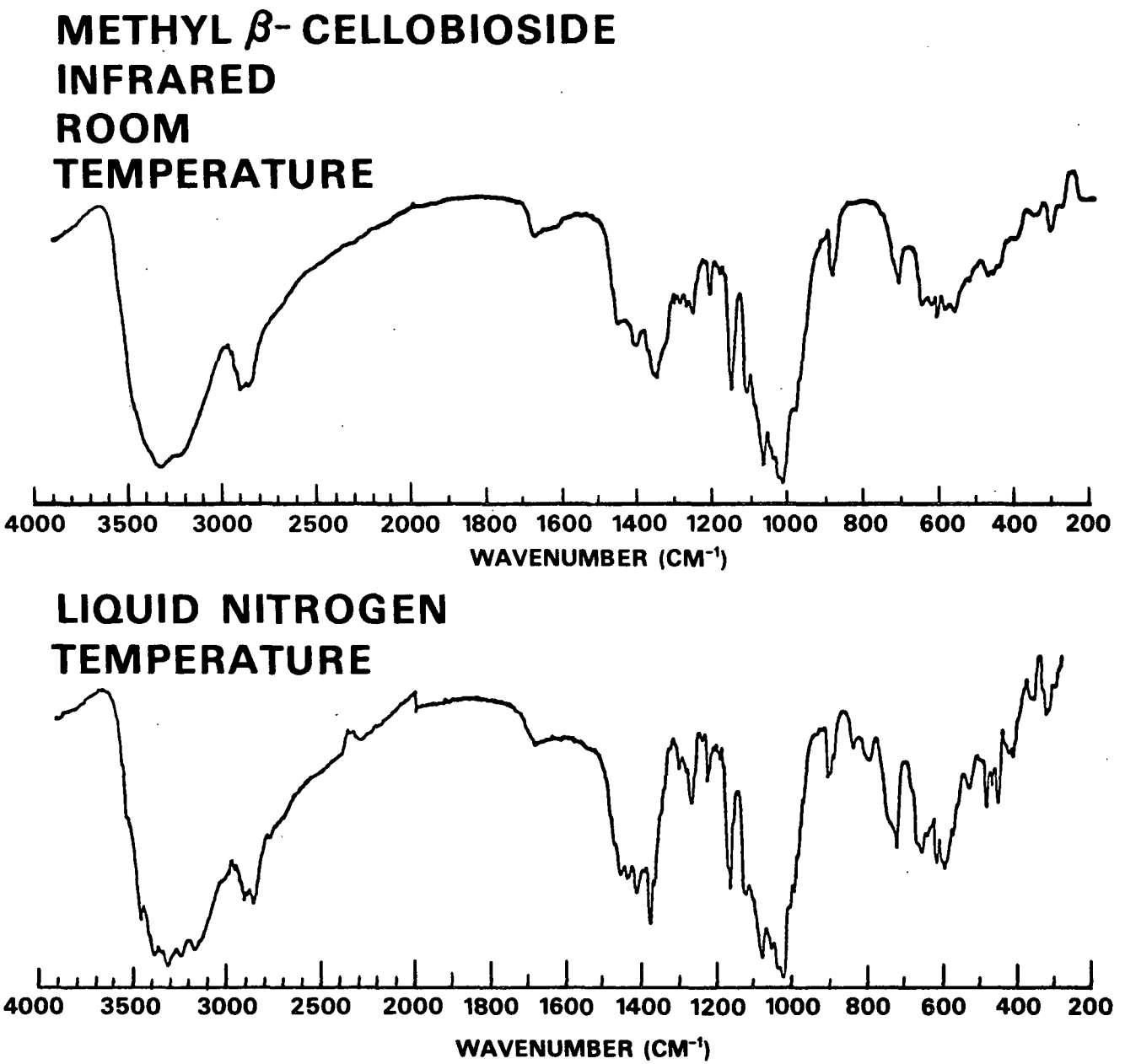


Figure 4. Infrared spectrum of methyl β-cellobioside.

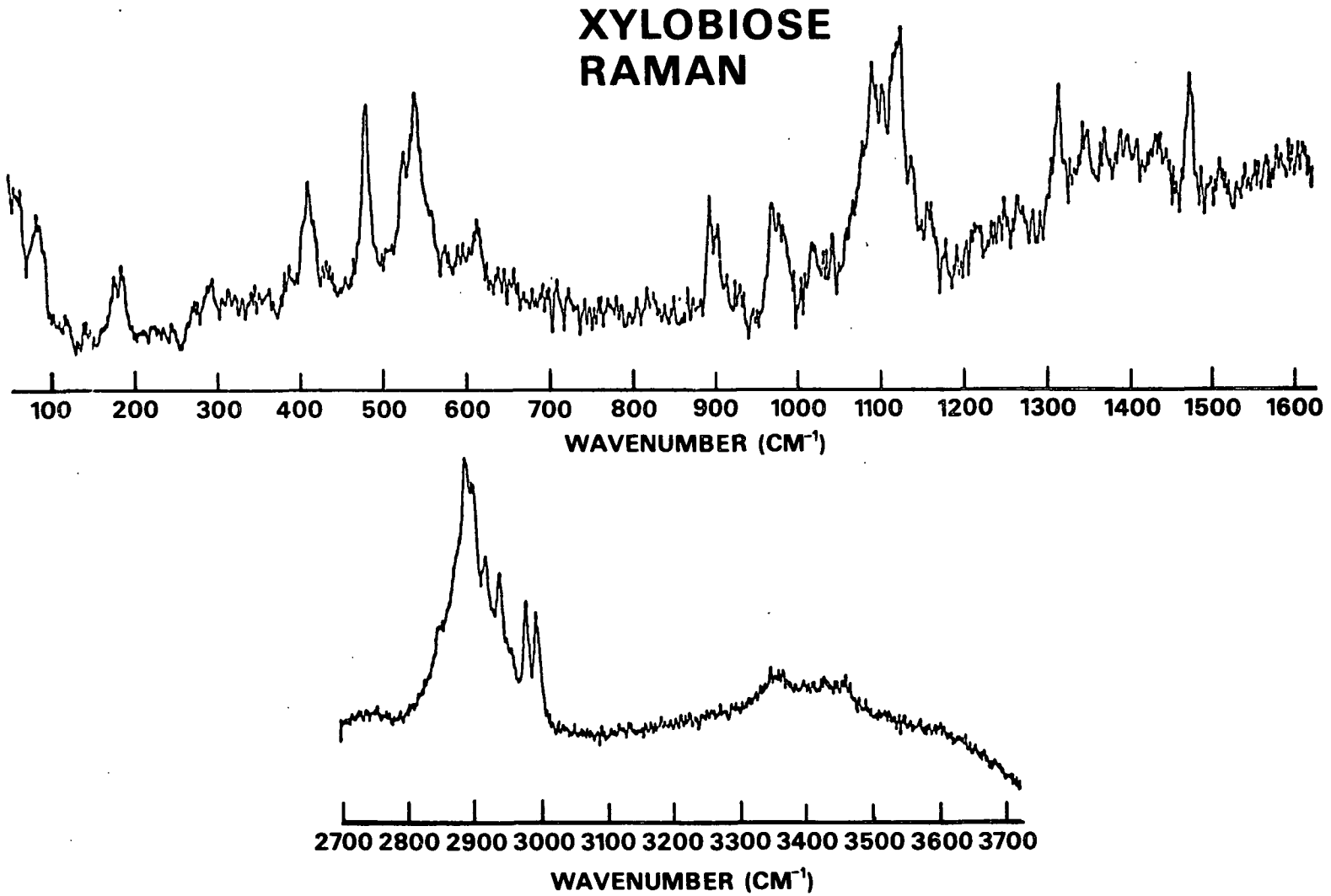


Figure 5. Raman spectrum of xylobiose.

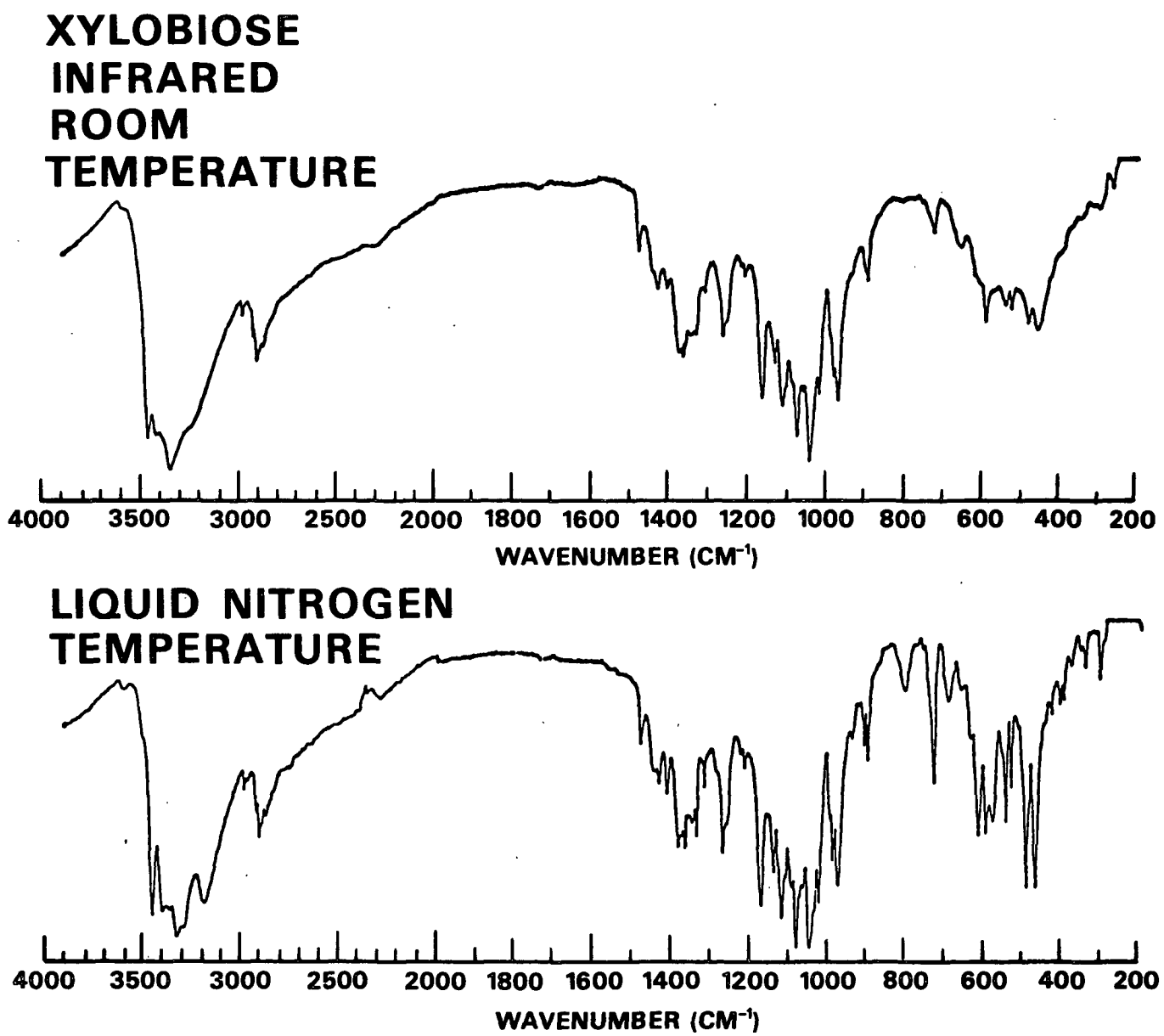


Figure 6. Infrared spectrum of xylobiose.

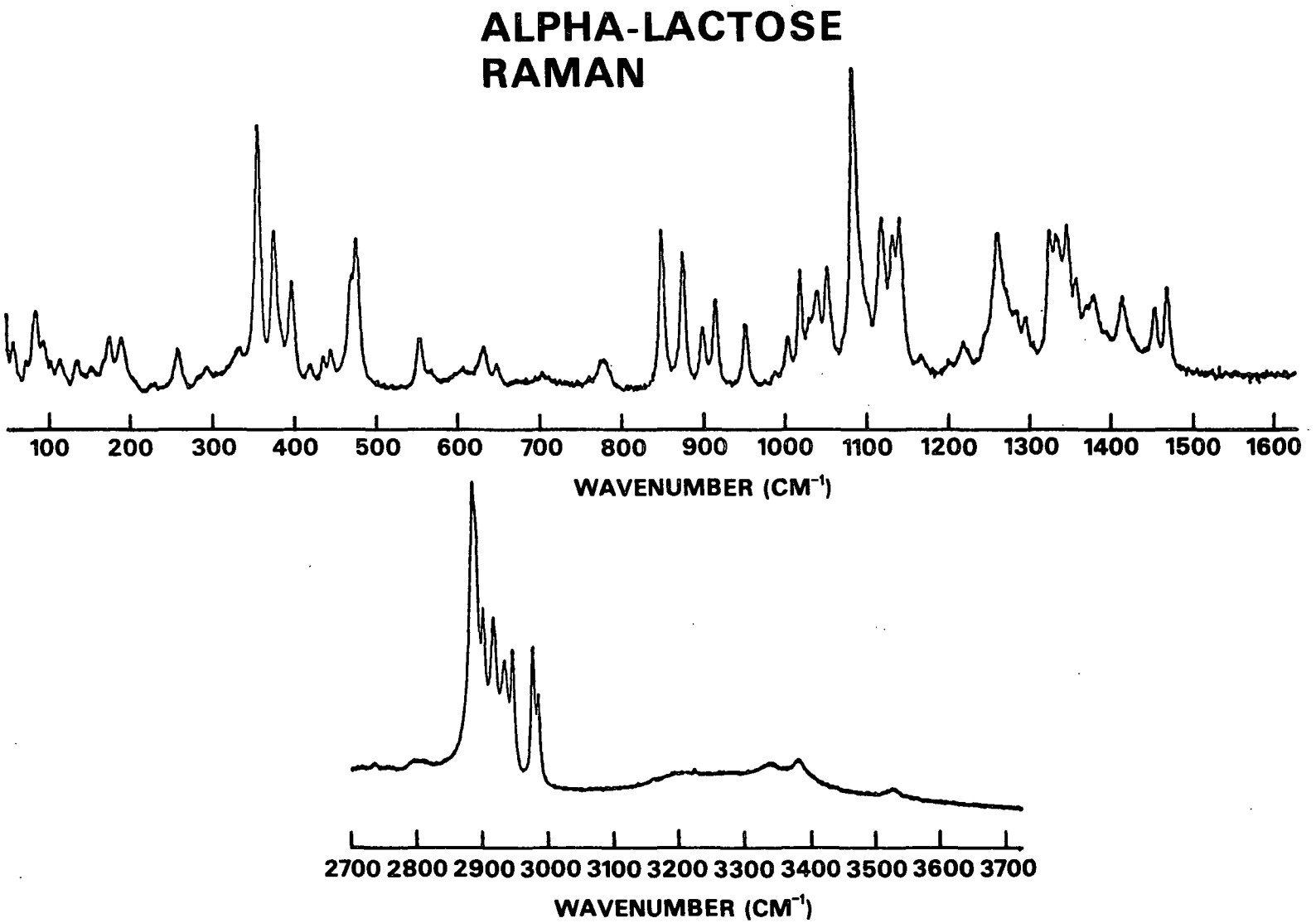


Figure 7. Raman spectrum of α-lactose.

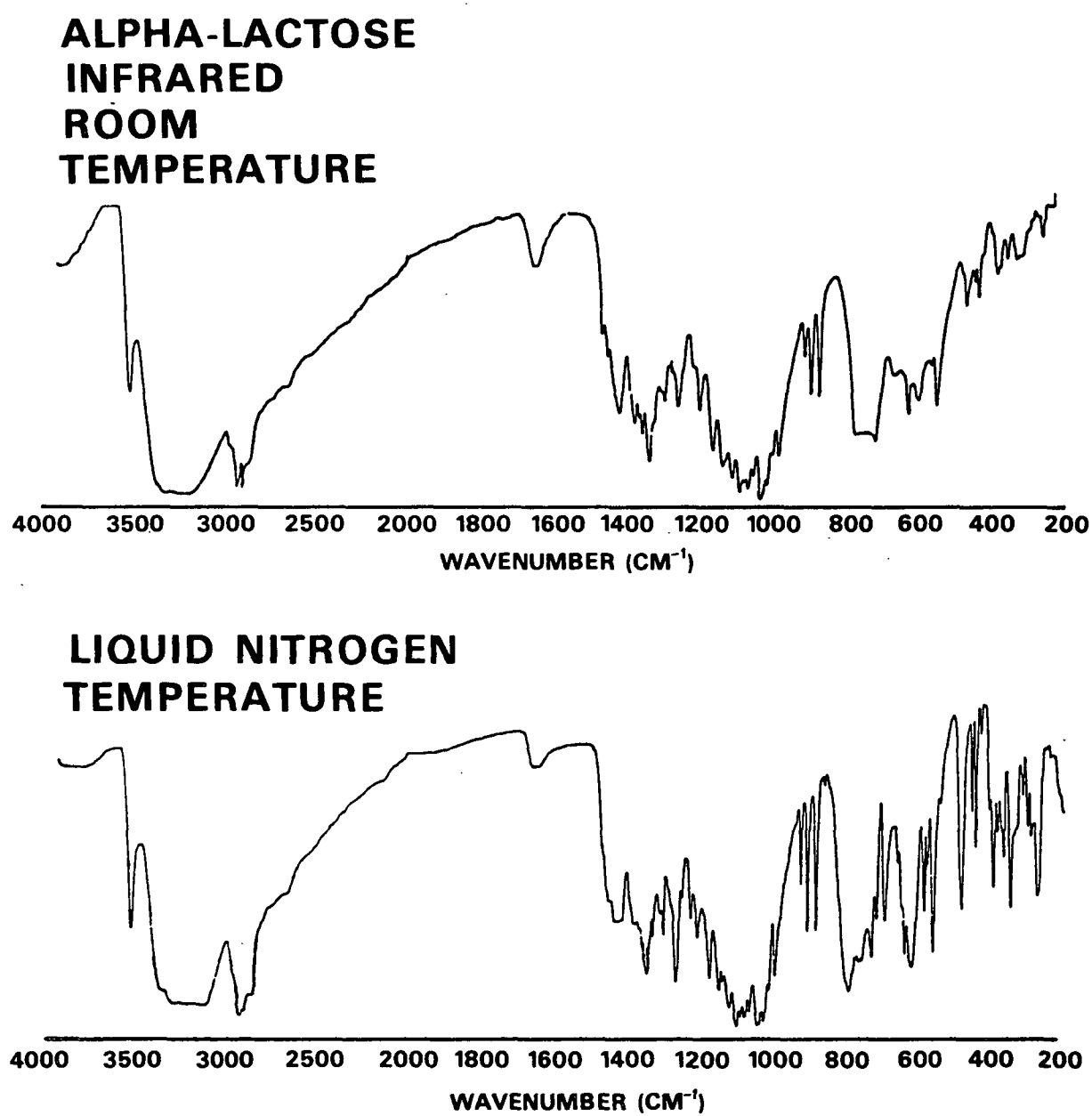


Figure 8. Infrared spectrum of α-lactose.

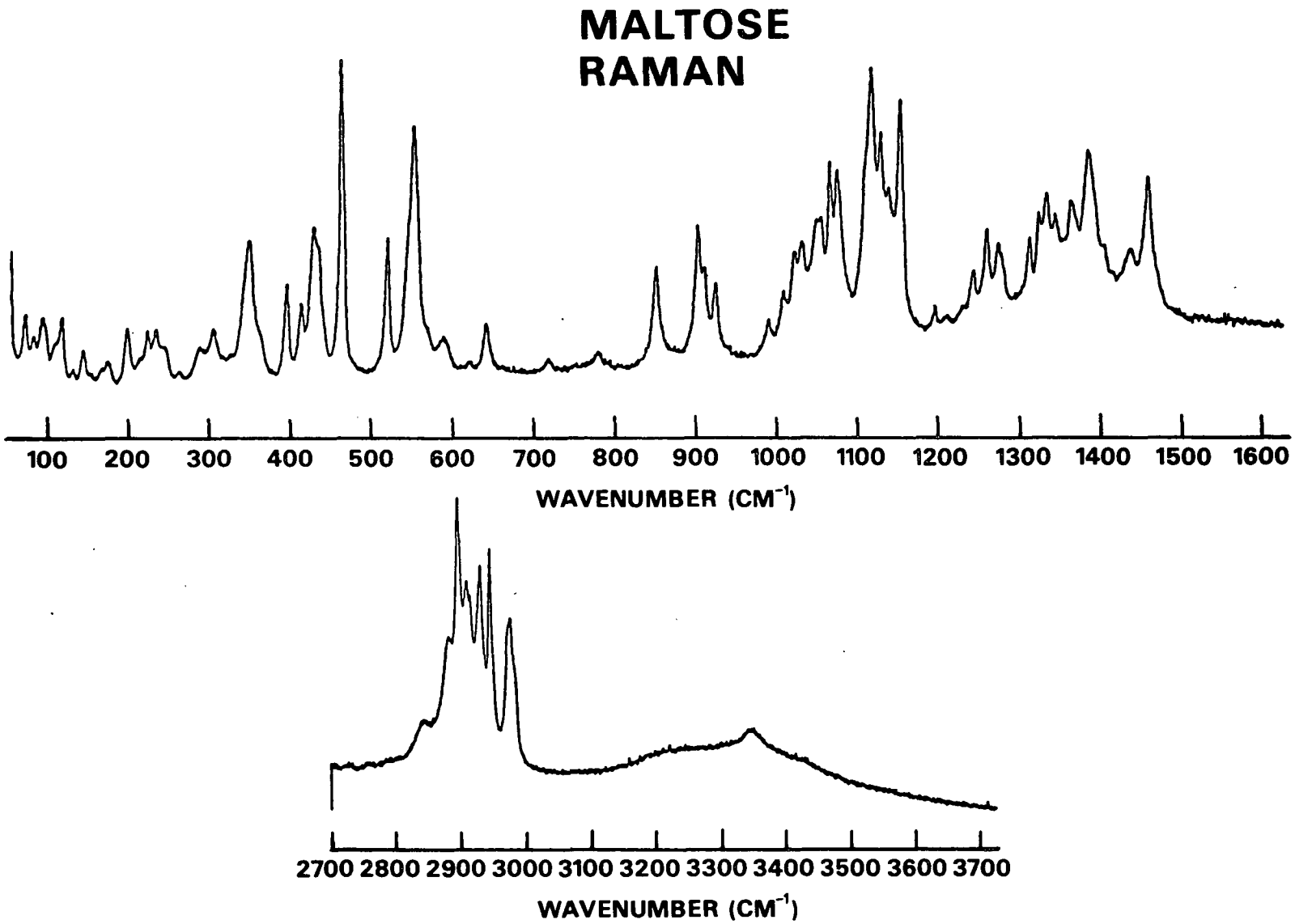


Figure 9. Raman spectrum of maltose.

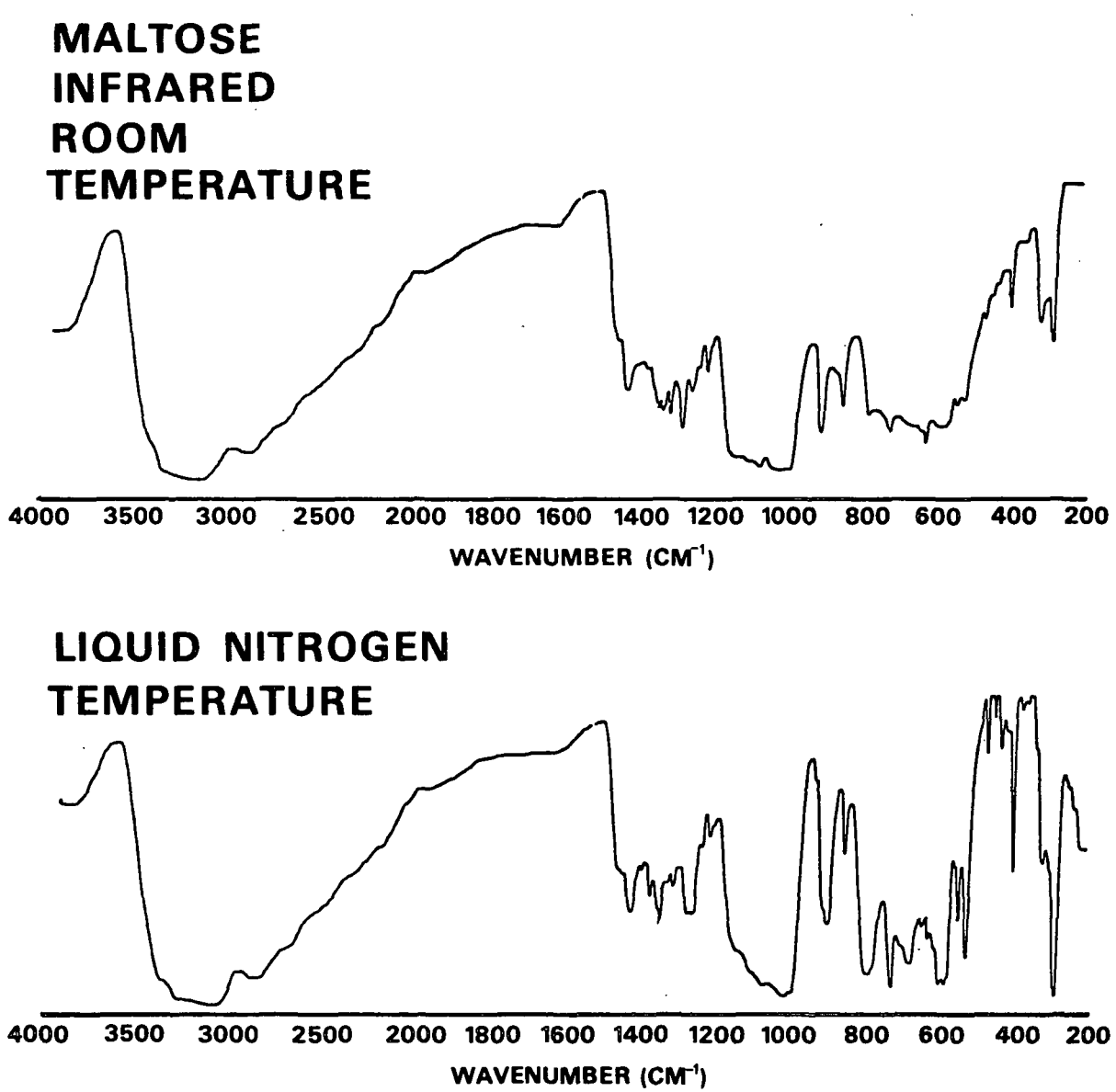


Figure 10. Infrared spectrum of maltose.

NORMAL COORDINATE ANALYSIS

Introduction

The logic involved in the normal coordinate analysis performed during the course of the thesis work is described below. Particular emphasis is placed on the choices of structures, internal coordinates, and force constants. The general flow of the computer programs with appropriate inputs and outputs is presented in Fig. 11. A detailed description of the normal coordinate analysis can be found in many texts[28-30] or in Appendix I. A copy of the programs used and some appropriate input data have been left in tape form at the Computer Center of The Institute of Paper Chemistry.

The computer programs are slightly modified versions of the normal coordinate software package written by Schachtschneider and Snyder[8,9] and some additional programs developed at IPC. It can be seen in Fig. 11 that the main inputs are structural information, internal coordinate definitions, and the "Z matrix" force constants. These inputs will be described below.

Structural Information

The structures used for the normal coordinate calculations of cellobiose, methyl β-cellobioside, and α-lactose were based on the reported crystal structures of these compounds.[18,31,32] The normal coordinate calculations assume the molecule to be totally isolated. The methanol in the unit cell of methyl β-cellobioside and the water in the unit cell of α-lactose were not included in the structures for the calculations to satisfy the isolated molecule approximation. The program CARTSET was used to determine bond lengths, bond angles, and dihedral angles from the fractional atomic coordinates for these compounds. This information formed the basis for the structures used in the calculations.

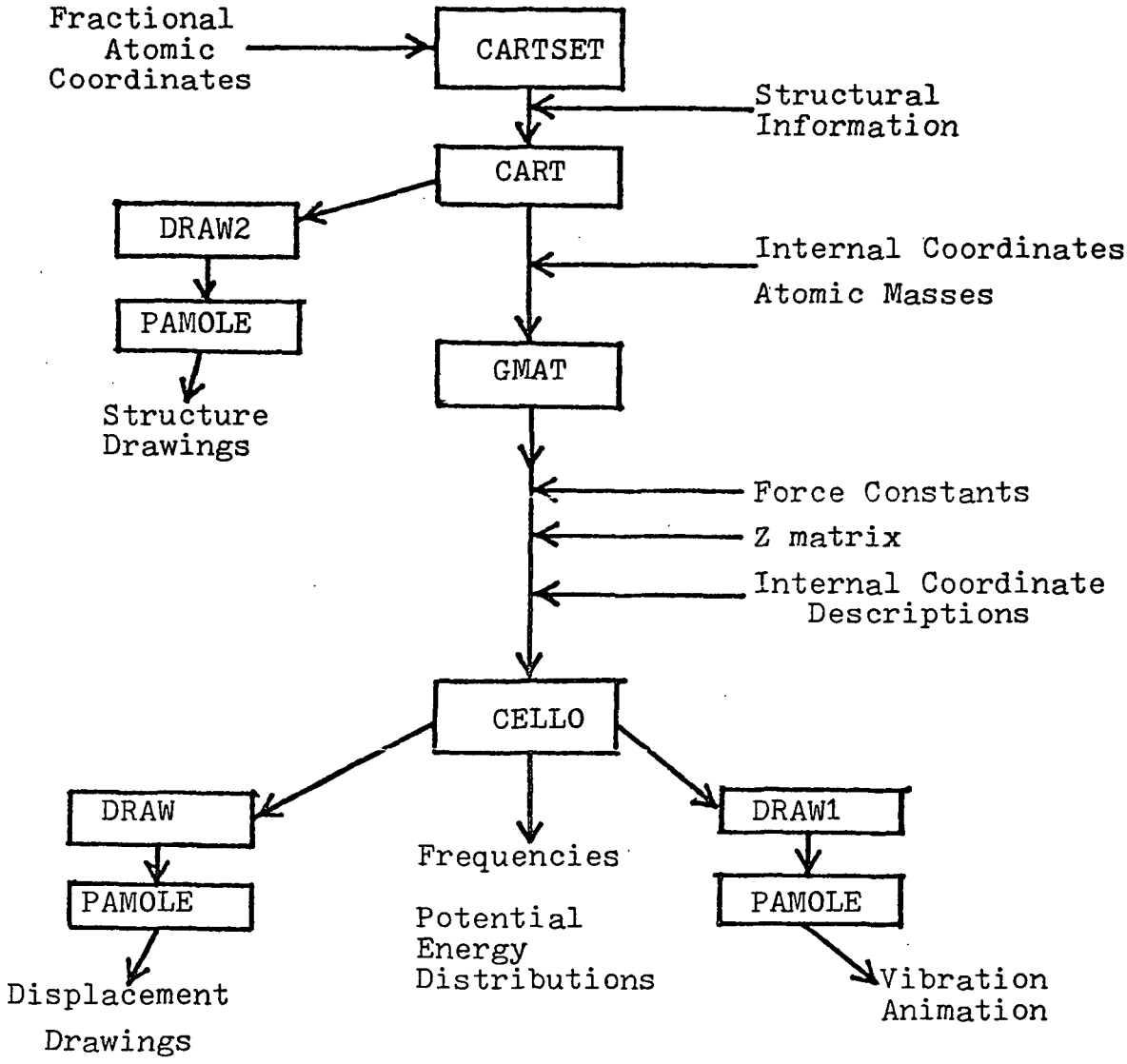


Figure 11. General flow diagram of computer programs.

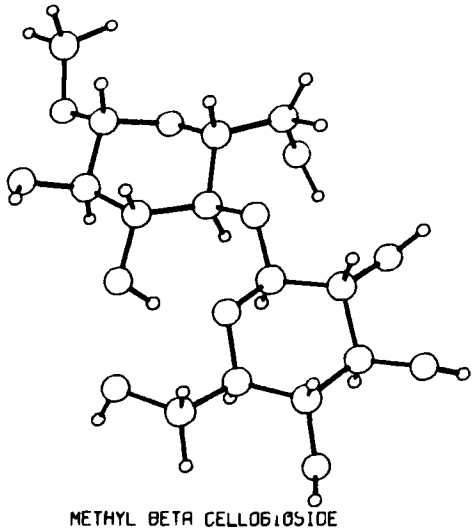
The only changes were made with regard to the hydrogen positions. The C-H bond lengths were set at 1.10 Å and the O-H bond lengths were set at 0.97 Å regardless of the values obtained from the fractional atomic coordinates. This was done because the x-ray crystal structures yield very short C-H and O-H bonds and to be consistent with previous work on NCH.[21-24] The dihedral angles defining the hydrogen positions within the hydroxymethylene units were set such that the carbon assumed an idealized tetrahedral configuration. The structures of

cellobiose, methyl β-cellobioside, and α-lactose are illustrated using "ball and stick" drawings in Fig. 12. The CARTSET and CART input data for these three disaccharides are included in the data on tape at the Computer Center.

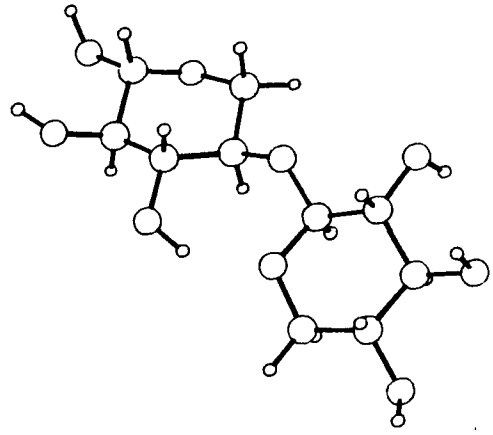
Figure 12 also included representations of the structures used for xylobiose and maltose. The structure of xylobiose was obtained by taking the above cellobiose structure, removing the hydroxymethylene units and assuming that the C5's were tetrahedral. This was done because there is no reported crystal structure for xylobiose. A reported crystal structure could not be found for maltose. However, there is a reported crystal structure for methyl β-maltopyranoside.[33] The structure used for maltose was obtained by taking this methyl β-maltopyranoside structure, removing the methyl group and making the same changes regarding the hydrogen positions as was made for cellobiose.

Internal Coordinate Definitions

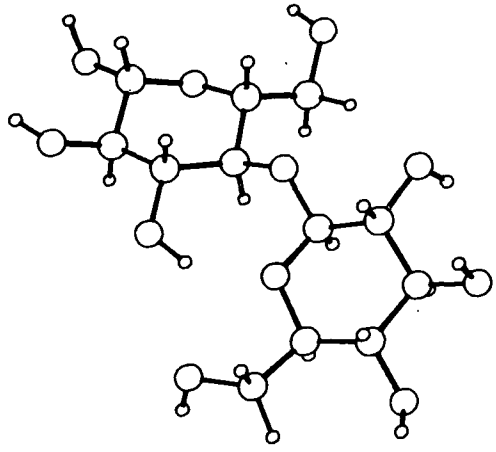
Imperative to the normal coordinate calculations is the selection of a set of internal coordinates which may be used to describe all motions of vibration for the molecule under consideration. The number of internal coordinates is 3N-6, where N is the number of atoms. Having more than 3N-6 internal coordinates "overdefines" the system and leads to redundant coordinates. However, having these redundant coordinates does not hinder the calculation of a unique transformation between cartesian to internal coordinates. The types of motion used in the normal coordinate calculations for the compounds under consideration consist of bond stretching, bond bending, and torsional motions. These motions, their determination, and the calculations associated with them are described in Appendix I. The bond stretching motions associated with the O-H and C-H bonds were not included in the calculations. It has been demonstrated[34] that the exclusion of these motions has only a slight effect on the calculations for



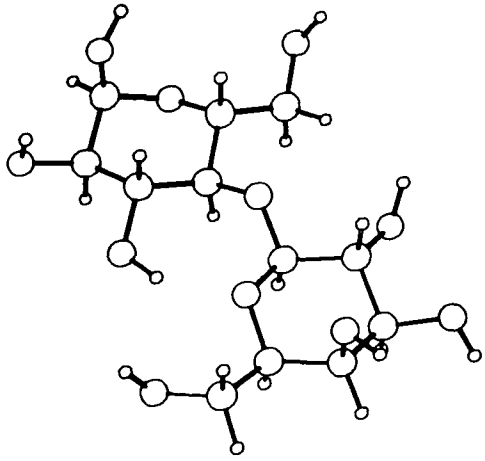
METHYL BETA CELLOBIOSIDE



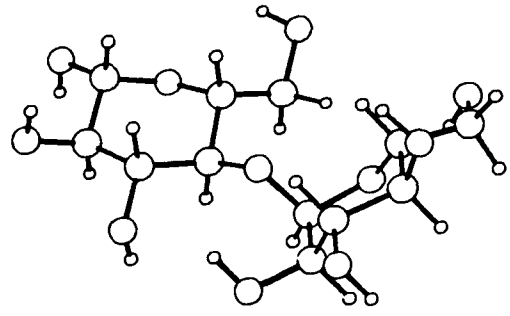
XYLOBIOSE STRUCTURE FOR CALCULATIONS



CELLOBIOSE CRYSTAL STRUCTURE



α - LACTOSE CRYSTAL STRUCTURE



MALTOSE STRUCTURE FOR CALCULATIONS

Figure 12. Structures used in calculations.

glucose. The number of internal coordinates can be obtained by taking 3N-6, adding all of the redundancies and subtracting the number of coordinates not included in the calculation. For example, the number of internal coordinates for cellobiose is 131. This is equal to 129 (3N-6) plus 24 (redundancies in each ring plus 1 redundancy at each carbon atom) minus 22 (14 C-H stretched and 8 O-H stretches).

Force Constant Selection

The potential energy for the calculated vibration of a molecule is described by a set of force constants for the internal coordinates and selected interactions between the internal coordinates. It is necessary to select force constants well suited to the molecules being considered in order to accurately reproduce the vibrations of the molecules. The majority of the force constants used in this investigation come from the normal coordinate analysis of the glucopyranose.[24] In that study Wells systematically adjusted the force constants to obtain the optimum fit of the calculated frequencies to the observed frequencies for several glucopyranoses. The similarity between the glucopyranoses and the compounds studied indicates that these force constants should be well suited for their intended purpose.

The force constants are the coefficients in the potential energy expansion in terms of internal coordinates. The expression is presented, in simplified terms, in Eq. (1). (See Appendix I for more details.)

$$2V = \sum_i \sum_j f_{ij}\, q_i\, q_j \qquad (1)$$

where V = potential energy

q = internal coordinate

f = force constant

Units of the force constants are therefore determined by the units of the internal coordinates. Table 1 presents the units of the internal coordinates and the diagonal force constants. The units of the off-diagonal, or interaction, constants may be determined from a combination of the units from the appropriate internal coordinates. The seemingly unusual units for the bending force constant results from requiring $f_{\alpha\alpha}\ \alpha^2$, where α is an angle, to have the same units of energy as $f_{rr}\ r^2$, where r is an extension of a bond. The interaction constants may be viewed as completing the potential energy expansion and an indication of the degree of interaction between the internal coordinates.

Table 1. Units of internal coordinates and diagonal force constants.

Internal Coordinate	Units of q_i	Units of f_{ii}
Bond stretching	Å	mdyne/Å
Bond bending	rad	mdyne Å/rad^2
Torsion	rad	mdyne/rad

The C-O-C bending motion located at the glycosidic linkage is not present in any of the hexoses studied by Wells. Therefore, a comparison of C-O-C bending constants was made to obtain a reasonable force constant for this motion. These values are contained in Table 2. From these data, a value of 1.315 mdyne Å/rad^2 was selected.

Table 2. Comparison of C-O-C bending constants.

Value (mdyne Å/rad^2)	Reference
1.260	Wells[24]
1.316	Edwards[23]
1.318	Pitzner[19]
1.313	Snyder and Zerbi,[11] also Cael,[14-16], Vasko[13], and Zhbankov[17]

Wells[24] assumed that all of the C-O stretching motions could be described by a single force constant. This reduced the number of force constants required to describe the vibrations. The good agreement which was obtained between the calculated and observed frequencies attests to the validity of this assumption for the glucopyranoses. This assumption was questioned for the disaccharides and the cellodextrins because of the addition of the C-O stretching motion at the glycosidic linkage(s). The C-O stretching constants were compared for a wide range of compounds. It was found that there was a high degree of correlation between the bond distance and the value of the force constant. This is consistent with the previous observations made by Becker[35] for a series of halides. The comparison is presented in Fig. 13 for C-O stretching force constant versus bond distance. The range of C-O bond lengths for cellobiose is also indicated in Fig. 13. The relationship between force constant and bond length was assumed to be linear in the region of C-O bond lengths found in cellobiose. The force constants for the C-O stretching motion for cellobiose were determined by the assumed linear relationship shown in Eq. (2). The values of the C-O stretching constants, C-O bond lengths, and references are given in Table 3.

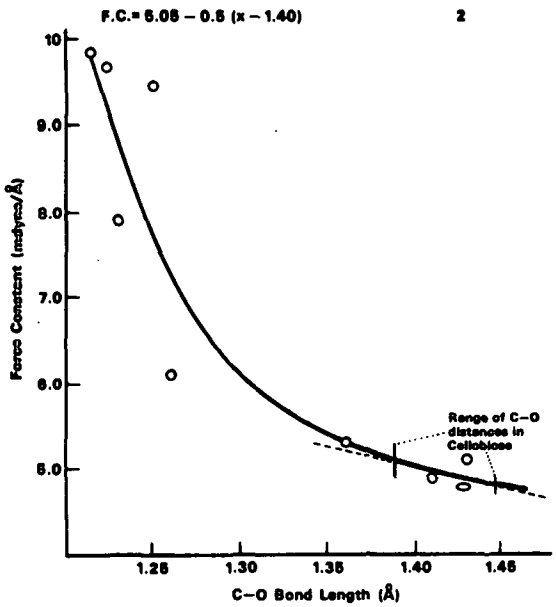


Figure 13. C-O stretching force constant (mdyne/Å) versus C-O distance (Å).

Table 3. C-O stretching force constants.

Force Constant, mdyne/Å	C-O Bond Length, Å	Source
8.911	1.216 (C=O)	Acetone, aldehyde, and formaldehyde by Cosee[36]
11.103	1.216 (C=O)	
9.2-9.7	1.25 (C=O)	Carboxylic acids[37]
5.3	1.36	Carboxylic acids[37]
6.11	1.26	Formic acid[38]
7.90	1.23	Cyclohexanone[39]
9.652	1.22	Ethers[11]
5.090	1.43	Glucose[24]
5.12	1.43 (approx.)	Hexoses[23]
5.151		1,5 Anhydropentitols[19]
5.103		
5.046		Alditols[22]
5.144		Inositols[25]

The force constants for the motions of the internal coordinates and the selected interactions are listed in Table 4. The heavy atoms, carbon and oxygen, have been grouped together for the interaction constants as they were for the glucopyranoses.[24] They are denoted by A, B, C, and D, or W, Z, Y, and Z for these interactions. The convention is that W-(X-Y)-Z denotes the bend-bend interaction for:

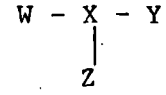


Table 4. Force constants.

Force Constant Number	Value	Description
	Stretching Constants (mdyne/Å)	
1	5.5000	O-H stretch
2	4.6500	C-H stretch
3	5.1336	C-O stretch
4	4.2124	C-C stretch
47	5.0600[a]	C-O stretch
48	5.0500[a]	C-O stretch
49	5.0300[a]	C-O stretch
50	5.0400[a]	C-O stretch
	Bending Constants (mdyne Å/rad^2)	
5	0.7700	C-O-H free
6	0.9500	C-O-H bonded
7	0.8660	O-C-H bend
8	0.5200	H-C-H bend
9	0.6643	C-C-H bend
10	1.0166	C-C-C bend
11	1.1114	O-C-O bend
12	1.3368	C-C-O bend
13	1.2631	C-O-C bend
51	1.3150	C-O-C bend

Table 4 (Continued). Force constants.

Force Constant Number	Value	Description
	Torsional Constants (mdyne/rad)	
14	0.0250	C-C-O-H torsion
15	0.1000	C-C torsion
16	0.0500	C-C ring torsion
	Stretch-Stretch Interactions (mdyne/Å)	
17	0.1063	C-C,C-C
18	0.3134	C-C,C-O
19	0.1203	C-O,C-O common C
20	0.1591	C-O,C-O common O
	Stretch-Bend Interactions (mdyne/rad)	
21	0.4808	C-C,C-C-C
22	0.4427	C-C,C-C-O
23	0.3733	C-O,C-O-C
24	0.5061	C-O,O-C-O
25	0.5069	C-O,C-C-O
26	0.3416	C-C,C-C-H
27	0.5151	C-O,O-C-H
28	0.4189	C-O,C-O-H

Table 4 (Continued). Force constants.

Force Constant Number	Value	Description
	Bend-Bend Interactions (mdyne Å/rad^2)	
29	0.1571	H-C-C-H trans
30	-0.0239	H-C-C-H gauche
31	0.0250	H-C-C-C trans
32	0.0204	H-C-C-C gauche
33	0.1240	H-C-C-O trans
34	-0.1895	H-C-C-O gauche
35	-0.0902	H-O-C-C trans
36	-0.1668	H-O-C-C gauche
37	0.1849	H-C-O-C trans
38	0.0243	H-C-O-C gauche
39	0.0401	H-O-C-H gauche
40	-0.0122	A-B-C-B trans (heavy)
41	0.0382	A-B-C-D gauche (heavy)
42	0.0761	W-(X-Y)-Z
43	-0.0841	H-(X-Y)-Z
44	0.0040	H-(X-Y)-H
45	0.0315	W-(X-H)-Z
46	-0.0030	H-(X-H)-Z

[a]As determined from Eq. (2).

COMPARISON OF CALCULATED AND OBSERVED FREQUENCIES

The calculated frequencies for the five disaccharides are compared to representations of their respective Raman spectra in Fig. 14 through 18. The representations consist of computer plots of the Raman spectra with values taken every 6.4 cm^{-1}. The general trends of the observed frequency distributions are reproduced to a high degree by the calculated frequencies. No assignment of calculated to observed frequencies was attempted. No force constants were refined. Therefore, it is not appropriate to discuss the average error and the degree of correspondence of individual calculated to observed frequencies. Figures 14 through 18 are presented to illustrate that the calculated frequencies reproduce the general trends of the observed frequencies for all five disaccharides.

INTERPRETATION OF THE VIBRATIONAL SPECTRA

An important result of the normal coordinate calculations for the disaccharides is the potential energy distribution for each calculated frequency. The potential energy distributions are the basis for the interpretation of the observed frequencies in terms of atomic motions. The potential energy distributions are calculated in terms of the relative contribution of the diagonal force constants in the solution of the normal coordinate calculations. The potential energy contributions for a particular frequency will be approximately 100% provided that the contributions from off-diagonal terms are not significant. The deviation of this sum from 100% is an indication of the contribution from off-diagonal terms. This contribution of the off-diagonal terms or interaction constants is an indication of the amount of coupling between the internal coordinates for a particular frequency. Therefore, the degree to which the sum of

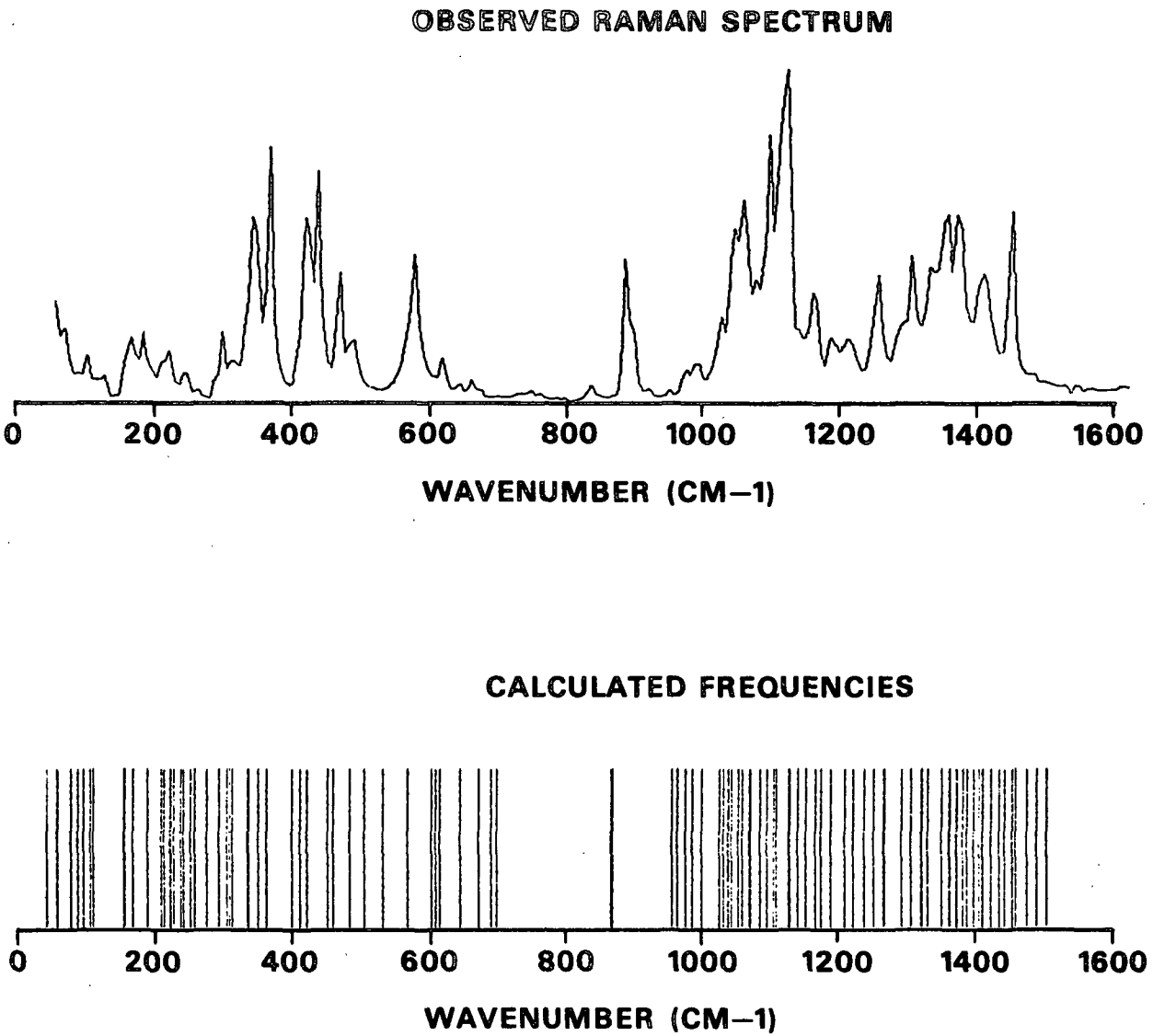


Figure 14. Calculated and observed spectra for cellobiose.

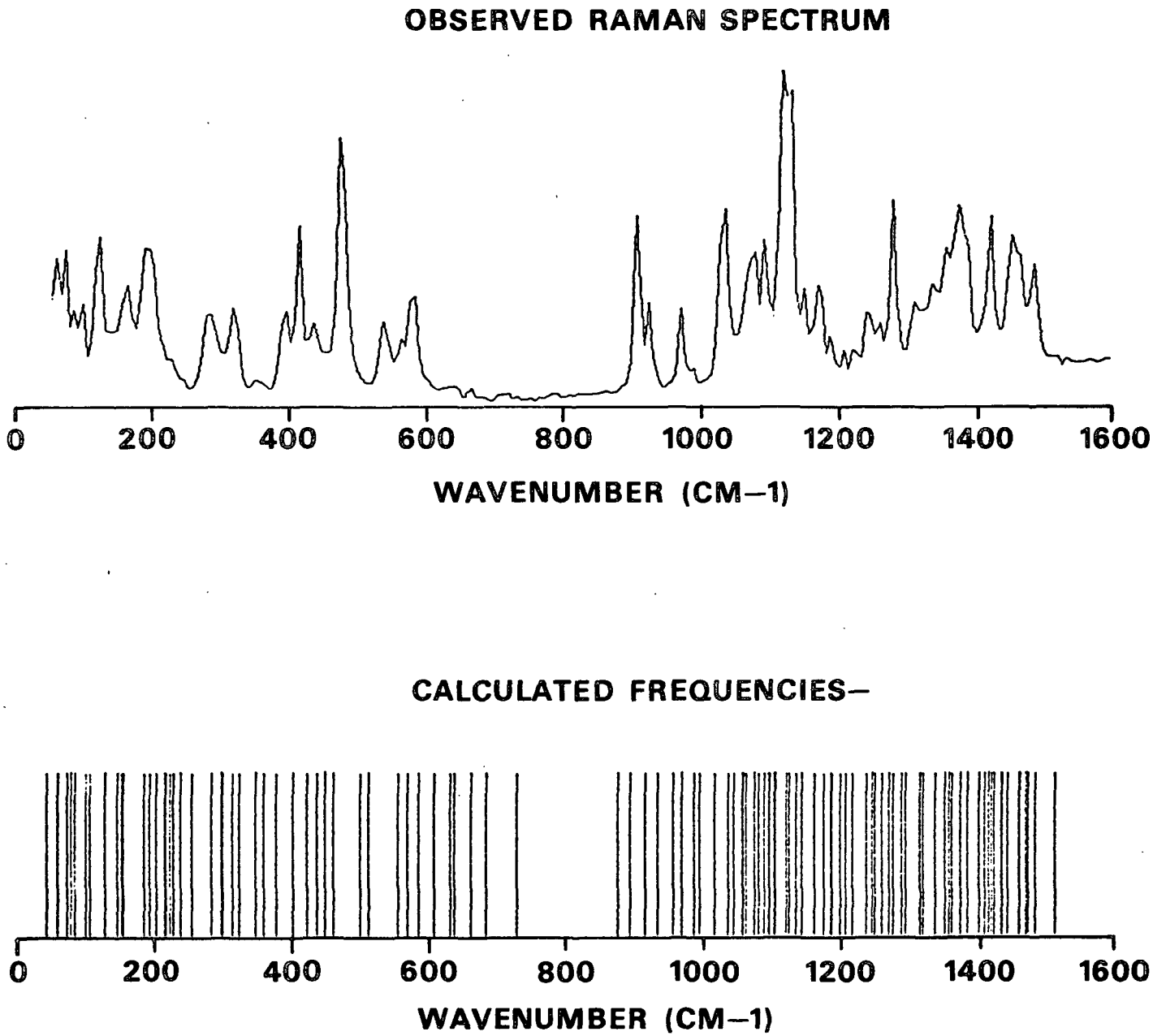


Figure 15. Comparison for methyl beta cellobioside.

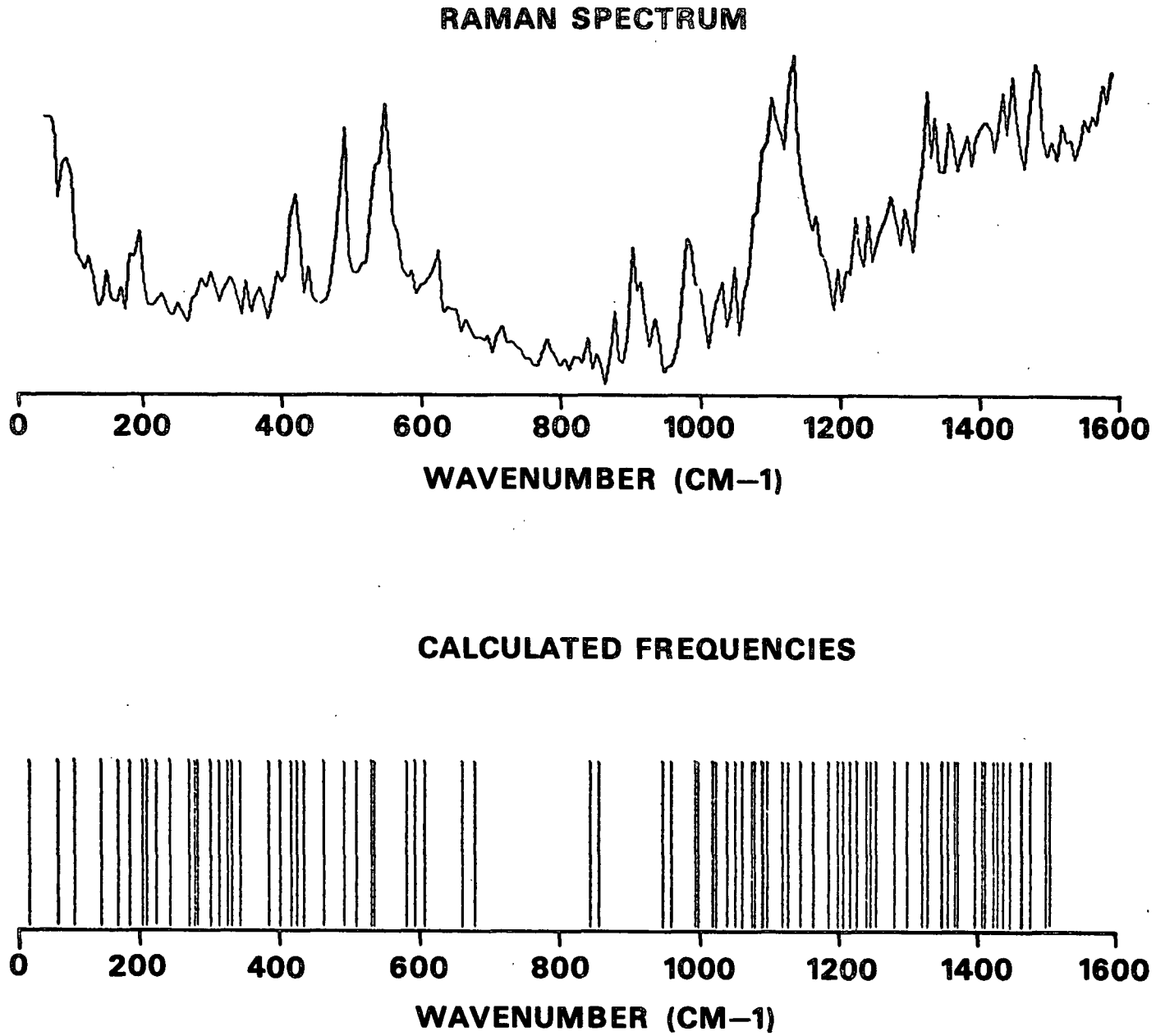


Figure 16. Calculated and observed frequencies for xylobiose.

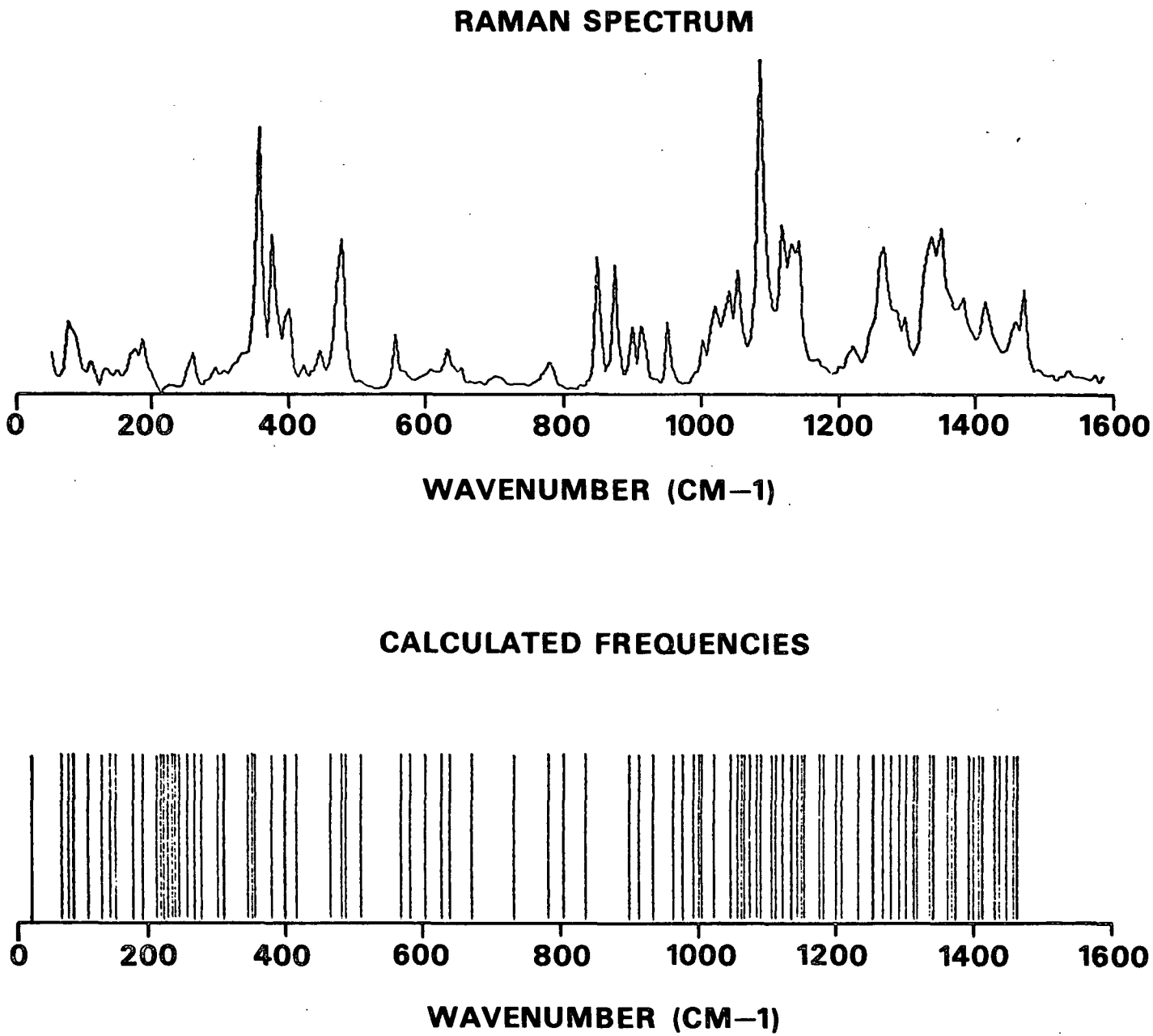


Figure 17. Calculated and observed frequencies for α-lactose.

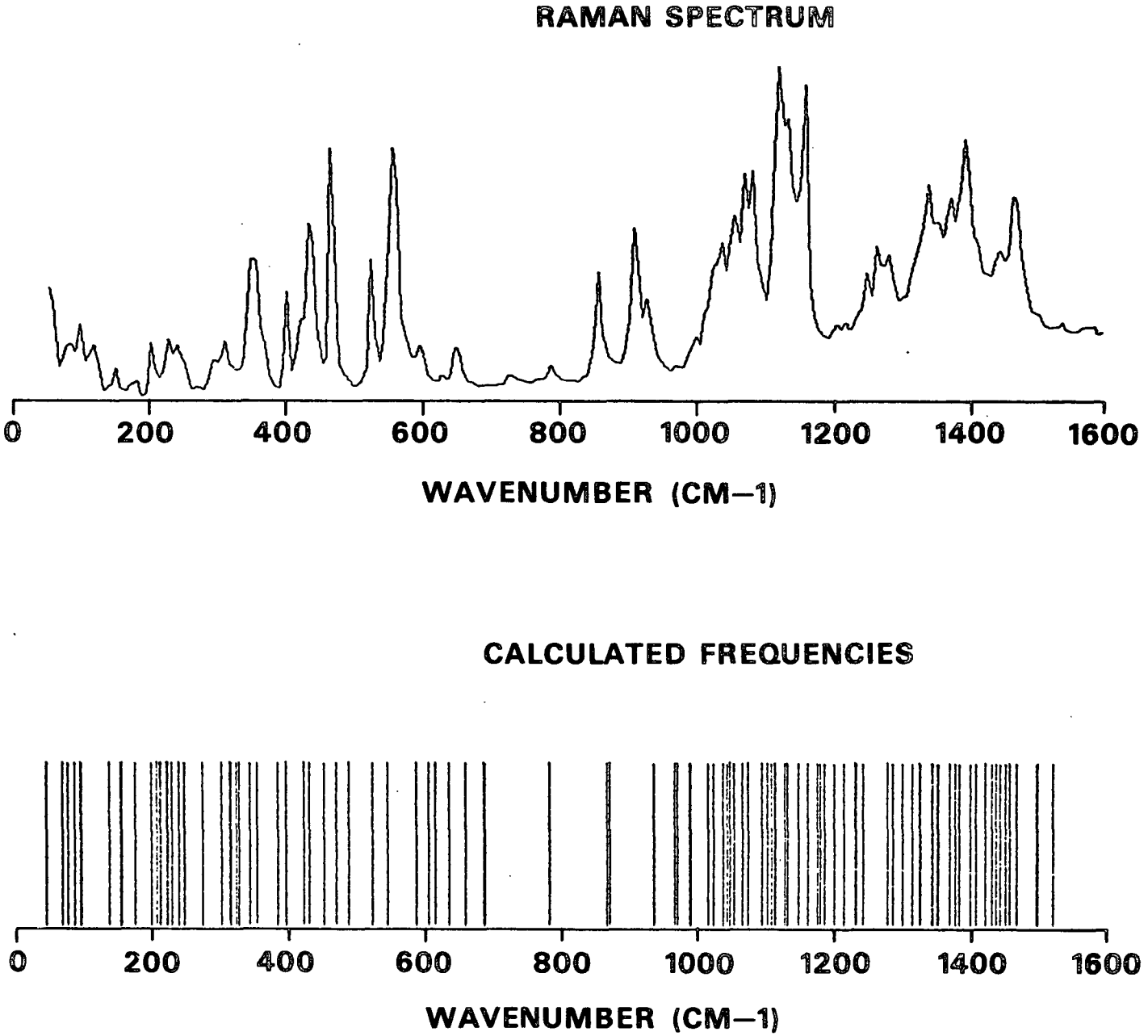


Figure 18. Calculated and observed frequencies for maltose.

the potential energy contributions deviate from 100% is an indication of the degree of coupling for that individual frequency.

Figure 19 provides a graphical summary of the potential energy contributions for the disaccharides investigated. The upper section described the normal modes in the conventional group frequency terminology and the lower area shows the primary ranges of potential energy contributions. These results agree with the group frequency literature[1-7] and, with only two exceptions, agree with the previously published carbohydrate calculations.[13-25] The first exception is that the C-O-H bending motions are calculated at slightly higher wavenumbers in this study than some previous studies[13-23] but in agreement with Wells' study of the hexoses.[24] Secondly, a large methine deformation localized at the hydroxymethylene (C6) positions was found at approximately 890 cm^{-1} in this study. This was reported for the normal coordinate analysis of cellobiose[17] but not in any of the other studies.[13-16,19-24]

The vibrational spectra of the disaccharides may be divided into six sections for the purpose of discussion. Regions II through VI are indicated in Fig. 19. Length prohibits the discussion of each individual band. The region 3500-2650 cm^{-1} encompasses the O-H and C-H stretching modes which were not included in the NCA. The methine (C-C-H and O-C-H), hydroxyl (C-O-H), and methylene (H-C-H) bending motions dominate the region 1500-1150 cm^{-1}. The contributions to the third region, 1150-850 cm^{-1}, are mainly the C-C and C-O stretching motions with the methine motions at the hydroxymethylene positions dominating the frequencies 890-850 cm^{-1}. The region 850-400 cm^{-1} is composed of heavy atom bending (C-C-O, O-C-O, and C-O-C) and heavy atom stretching motions. The fifth region, 400-300 cm^{-1}, is primarily composed of C-C-O and C-C-C bending

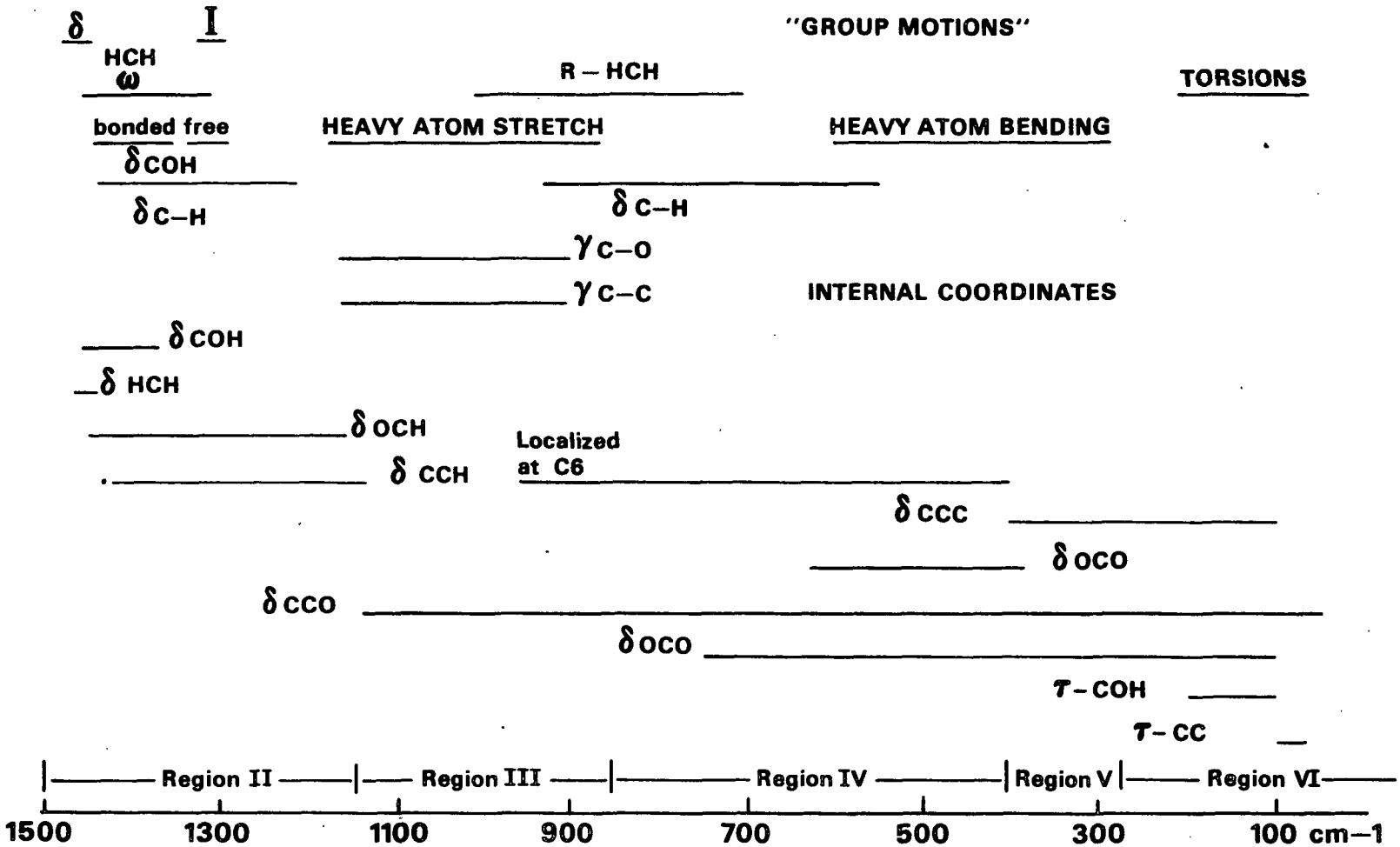


Figure 19. General potential energy distributions for the disaccharides.

with some C-C and C-O stretching motions. Below 300 cm^{-1} the calculated frequencies are dominated by the torsional motions combined with some contribution from the heavy atom bending. This division into six regions is not meant to constitute a rigorous categorization of the different types of atomic motions. The majority of the calculated frequencies are composed of many types of motion and the complexity of the motions associated with these frequencies will become apparent in the following discussion.

Figures 20 through 22 are computer drawings of the potential energy contributions for the calculated frequencies between 1600 cm^{-1} and 50 cm^{-1}. The data for these drawings are the relative contributions of the diagonal force constants to each calculated frequency. These are referred to as the potential energy distributions (or PEDs). As was discussed earlier, the sum of the PEDs may deviate from 100% and the amount of this deviation can be used as a measure of the relative coupling between the internal coordinates. Figures 20 through 22 are the basis for the drawings which will be used in the following discussion of the individual regions.

Region I: 3500-2650 cm^{-1}

The region 3500-2650 cm^{-1} contains the bands in the vibrational spectra arising from the O-H and C-H stretching motions. As discussed earlier, these motions were not included in the normal coordinate calculations. This exclusion is comparable to assigning infinite force constants to these stretching motions and disregarding any interactions which may occur with the included motions. The effect of not including these motions for glucose was investigated by Wells.[34] He found that the exclusion had only a minimal effect on the calculated frequencies.

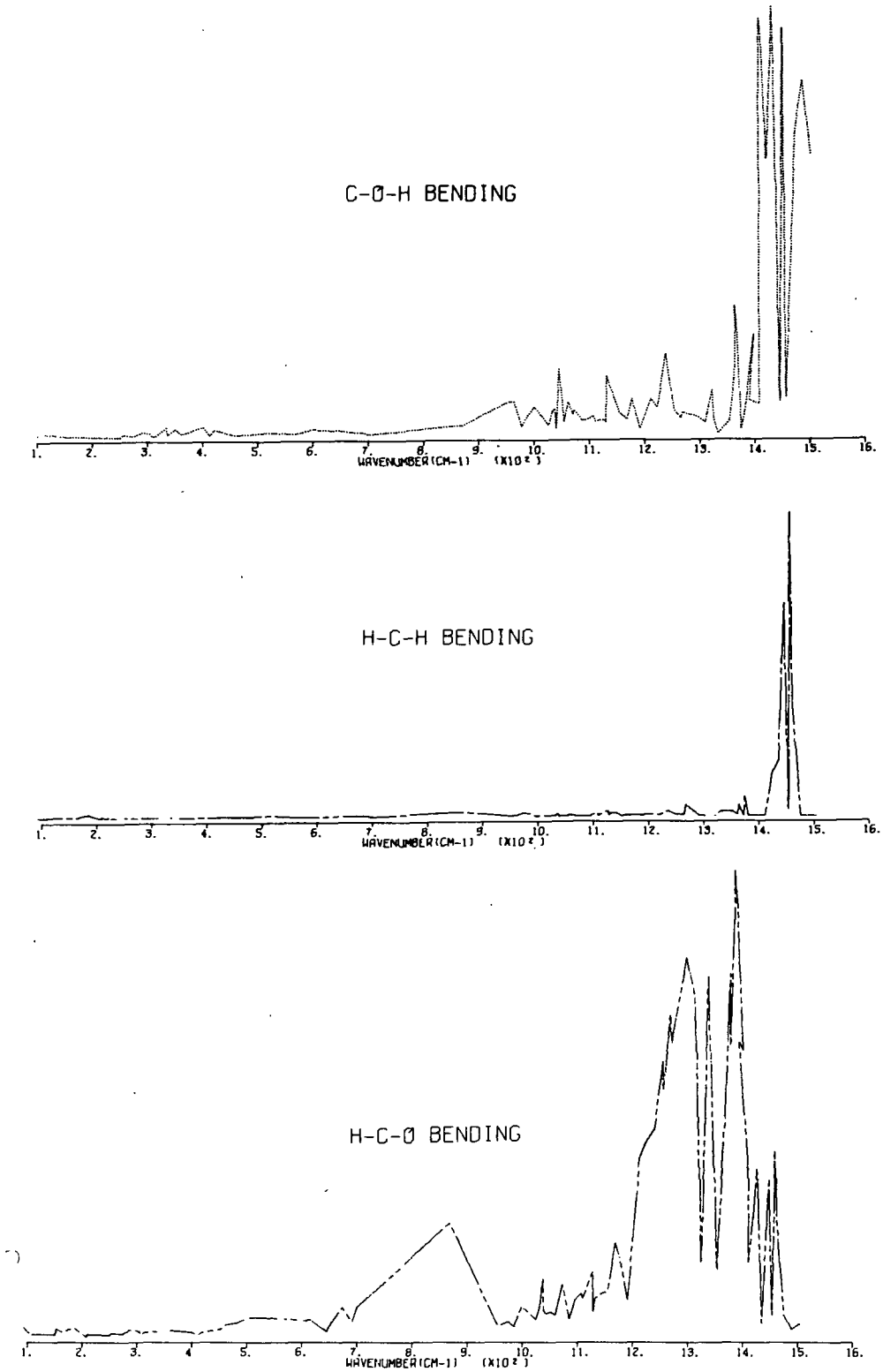


Figure 20. Potential energy contributions for C-O-H, H-C-H, and H-C-O bending motions.

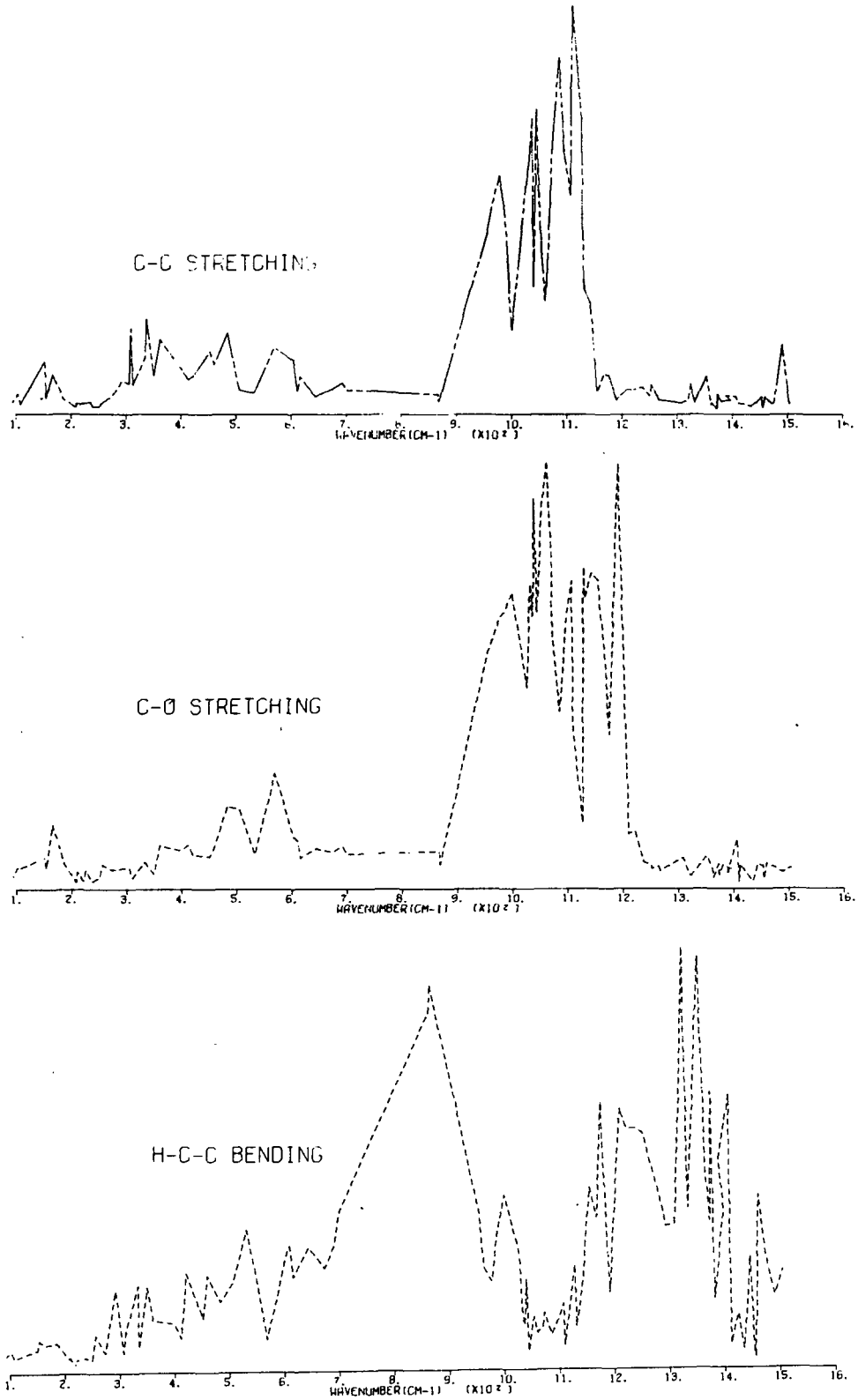


Figure 21. Potential energy contribution for H-C-C bending, C-C stretching and C-O stretching motions.

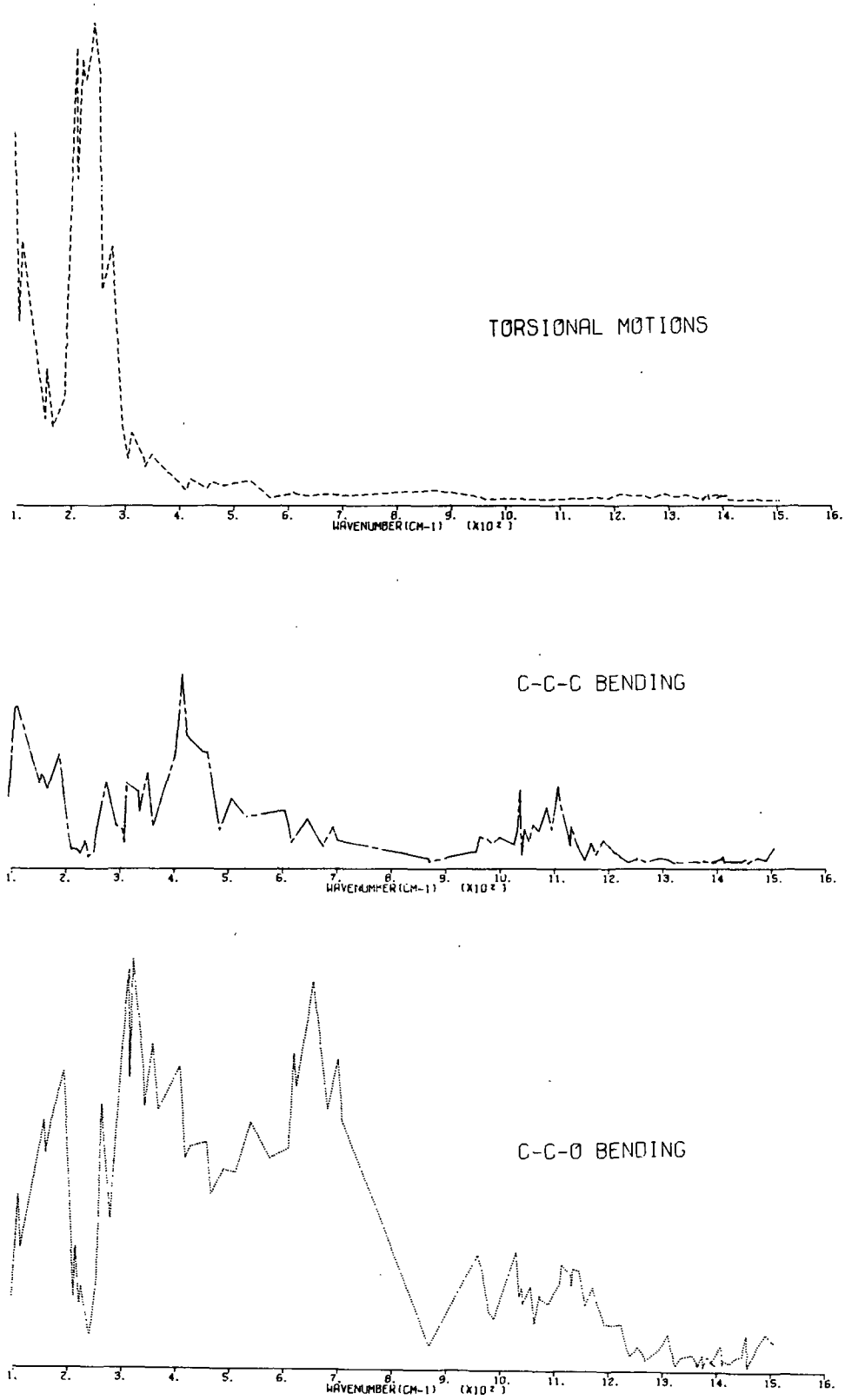


Figure 22. Potential energy contributions for torsional, C-C-C bending and C-C-O bending motions.

The O-H stretching modes for the disaccharides occur between 3500 and 3150 cm^{-1}. The bands in this region are affected by intermolecular and intramolecular hydrogen bonding and crystal effects. These bands are very intense in the infrared and weak to moderate intensity in the Raman. It is beyond the scope of this work to provide a detailed interpretation of these frequencies for the disaccharides.

The C-H stretching modes occur in the region 2970-2880 cm^{-1} and were observed in both the infrared and the Raman spectra of all the disaccharides. These bands are very intense in the Raman, often being the most intense bands in a given spectrum. In contrast, they are only of moderate intensity in the infrared.

Region II: 1500-1150 cm^{-1}

The potential energy distributions show this region to be dominated by the methylene (H-C-H), hydroxyl (C-O-H), and methine (C-C-H and O-C-H) bending motions. The relative contributions of these motions are illustrated in Fig. 23 for this region. Again, Fig. 23 is based on the computer drawings shown in Fig. 20.

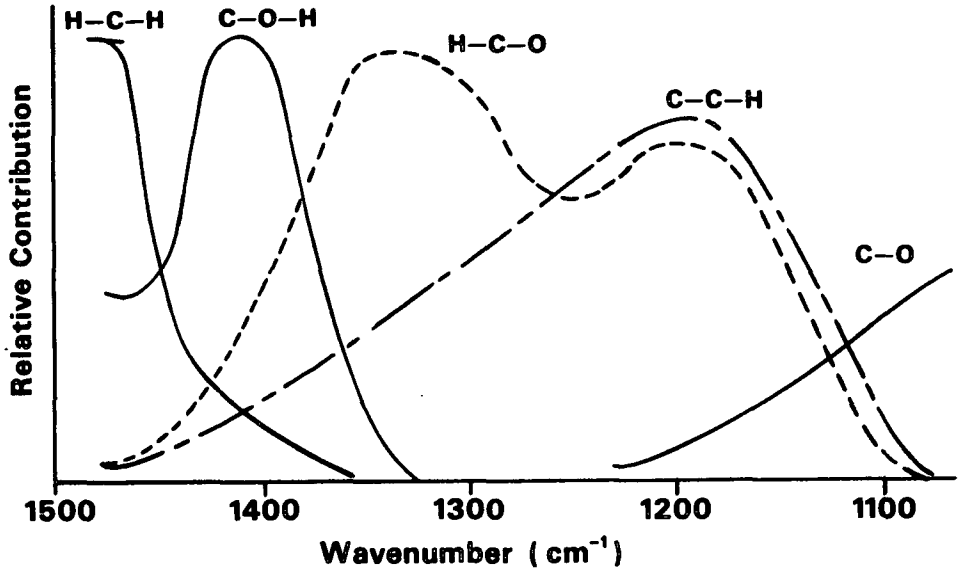


Figure 23. Major potential energy contributions for Region II.

The methylene bending motions dominate the higher frequency region 1465-1440 cm^{-1}. The displacements for the normal mode calculated at 1460 cm^{-1} for cellobiose are illustrated in Fig. 24. Drawings have been made for the various normal modes of all the disaccharides. All the disaccharides demonstrate similar motions for each region of their calculated spectra. Cellobiose was chosen as being representative to avoid confusion when comparing the displacements for the normal modes. The displacements in Fig. 24 are predominantly located at the hydroxymethylene units and may be termed H-C-H scissoring motions. The hydroxyl bending motions dominate the region 1440-1400 cm^{-1}. The region 1400-1300 cm^{-1} has major contributions from the O-C-H bending motions which then combine with the C-C-H bending motions to dominate the region 1300-1150 cm^{-1}. These motions are very complex but are composed primarily of the motions of the hydrogen atoms. Representative displacements are illustrated in Fig. 25 and 26 for the modes calculated at 1268 and 1222 cm^{-1}.

Region III: 1150-850 cm^{-1}

The potential energy contributions for Region III are illustrated in Fig. 27. It can be seen that the major contributions for this region are the C-C and C-O stretching motions with the methine bending motions at the hydroxymethylene positions dominating the region 890-850 cm^{-1}. The observed spectra for all the disaccharides show very intense bands in this region in both the infrared and the Raman. The calculated frequencies for this region are very complex but can be divided into two classes. The first class is illustrated in Fig. 28 with the displacements for the cellobiose frequency calculated at 1072 cm^{-1}. This class of motion contains significant motions in both rings. This indicates a high degree of coupling of motions between the two rings. The other class of normal mode calculated in this region has displacements localized within one of the two

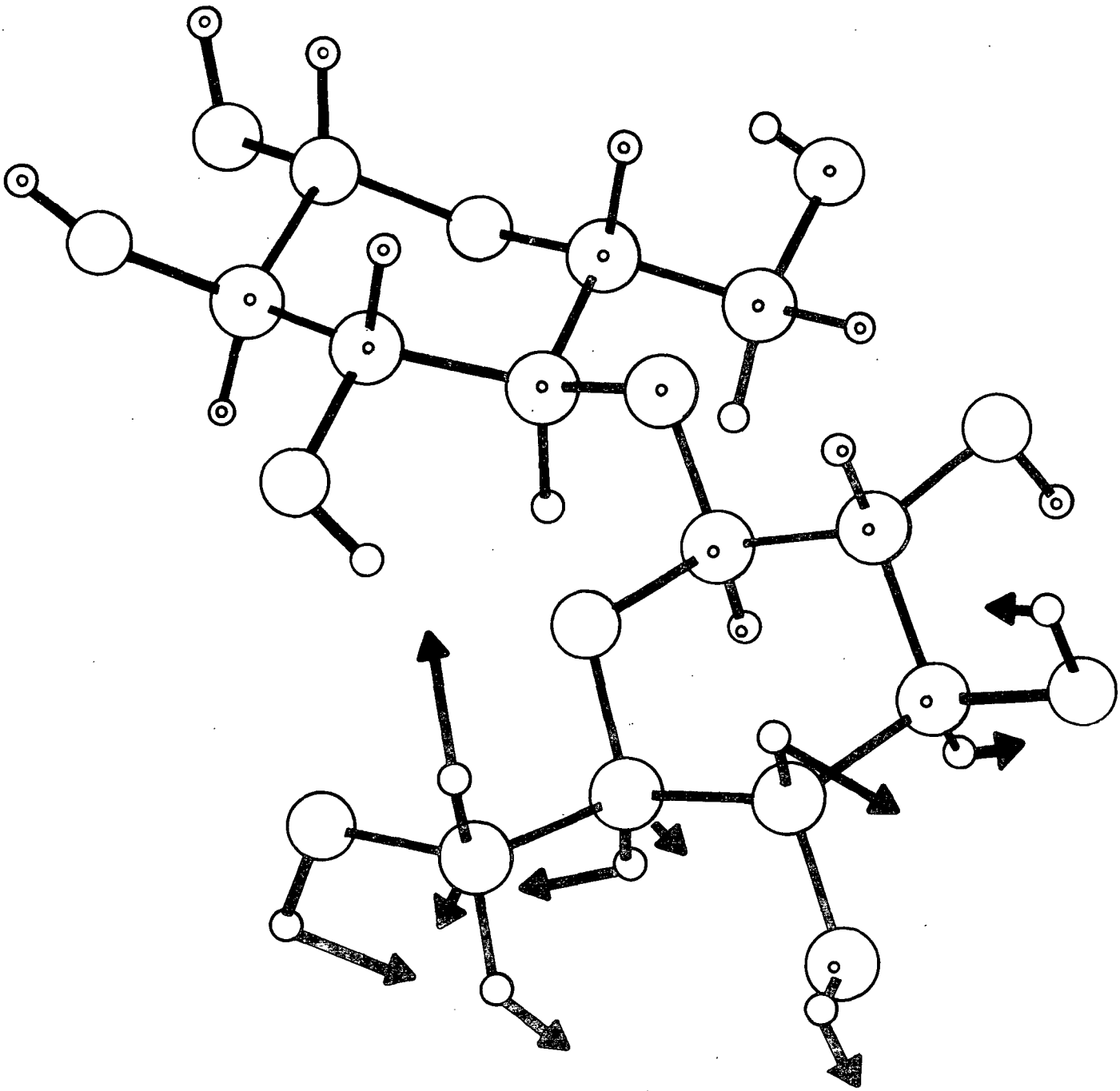


Figure 24. Displacements for the cellobiose calculated frequency at 1460 cm^{-1}.

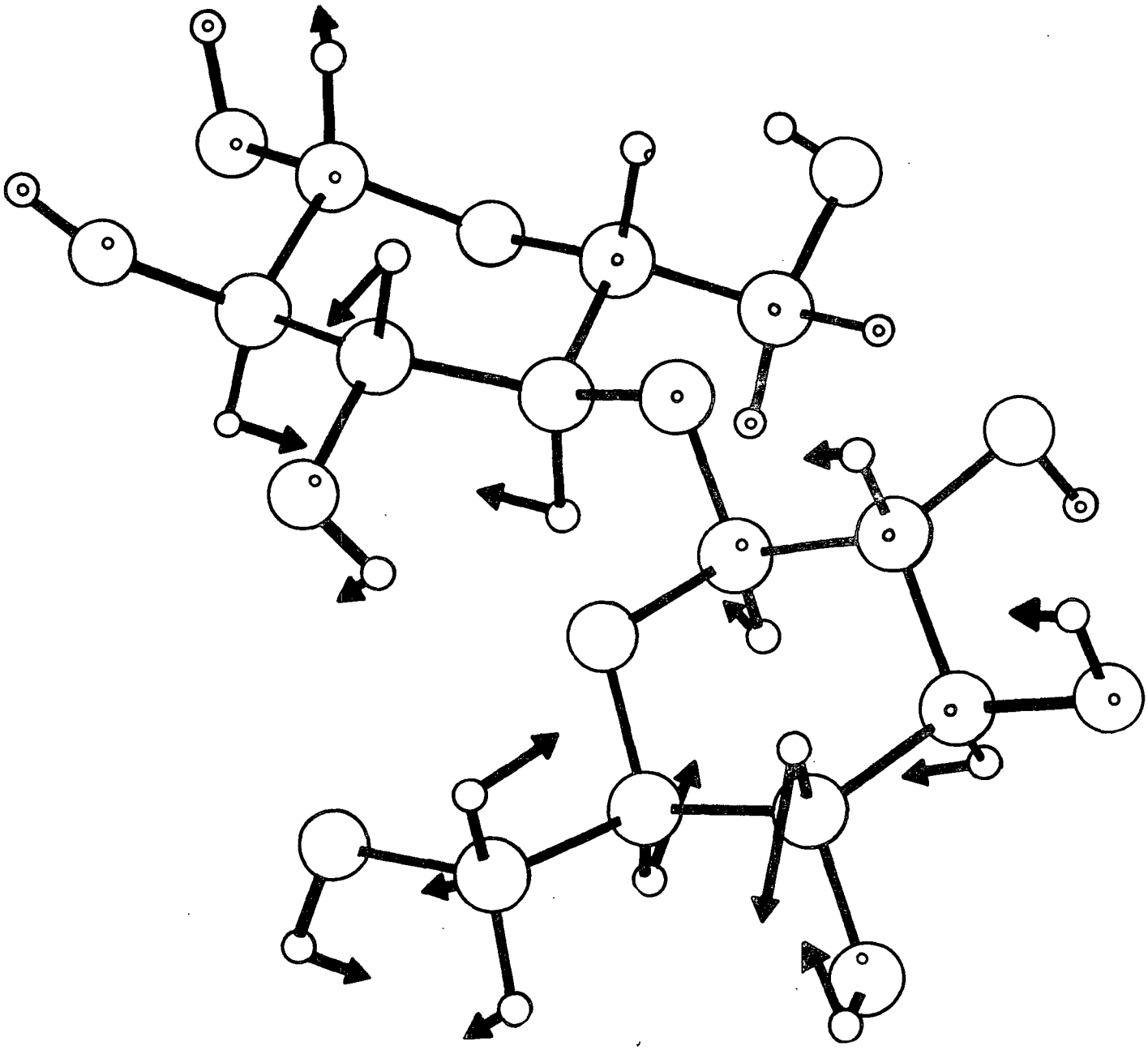


Figure 25. Displacements for the cellobiose calculated frequency at 1268 cm^{-1}.

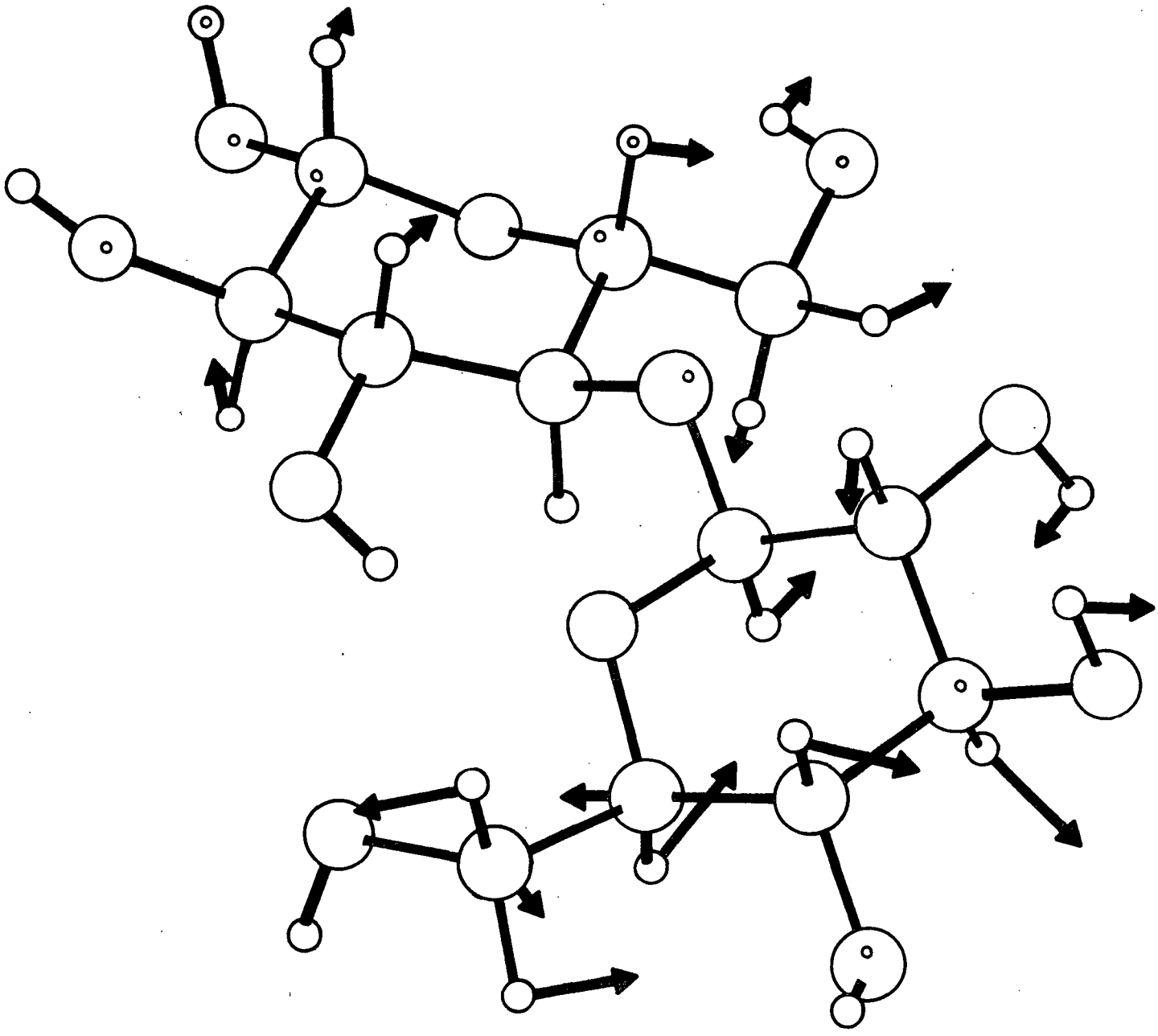


Figure 26. Displacements for the cellobiose calculated frequency at 1222 cm^{-1}.

rings. These frequencies have a complement having a very close value for its frequencies and a concentration of motion in the neighboring ring. A pair of these frequencies are illustrated in Fig. 29. In actuality, these frequencies should combine to form a single observable band which would then be composed of displacements in both rings. This also indicates a high degree of coupling of the motions between the rings which constitute the normal mode.

This coupling between the rings must occur through the glucosidic linkage. Any change in the glucosidic linkage, either its geometry or hybridization, should result in a change in this coupling. Therein lies the significance of the coupling, as a tool for the interpretation of frequency shifts and intensity changes in this area observed when cellulose is chemically and physically treated.

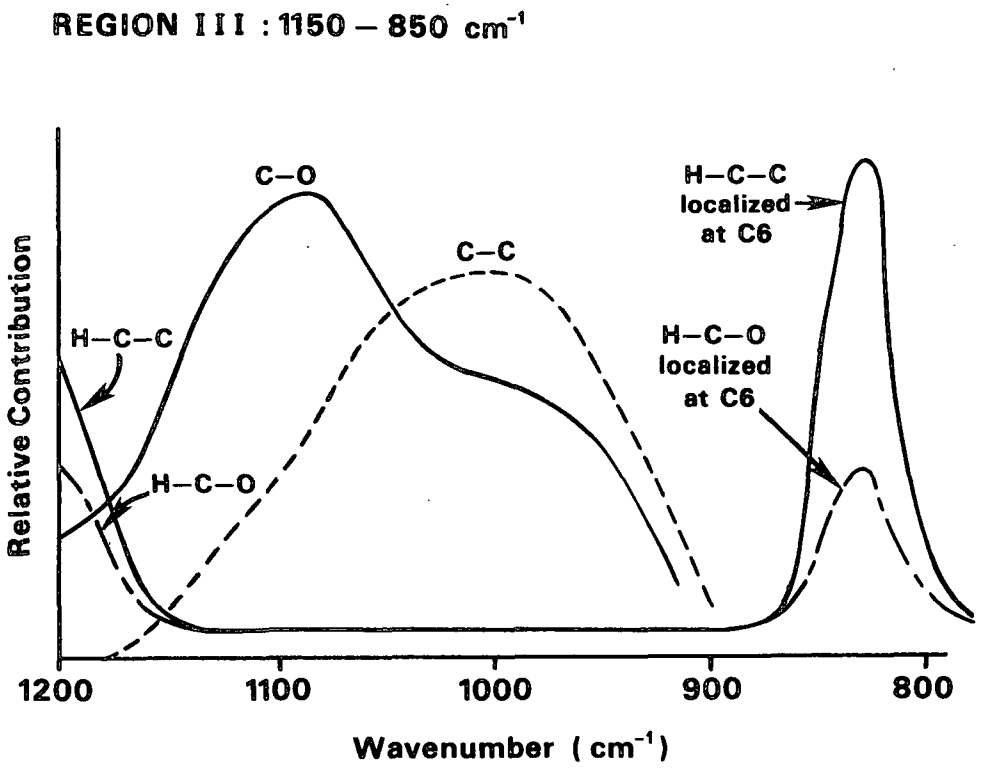


Figure 27. Relative potential energy contributions for Region III.

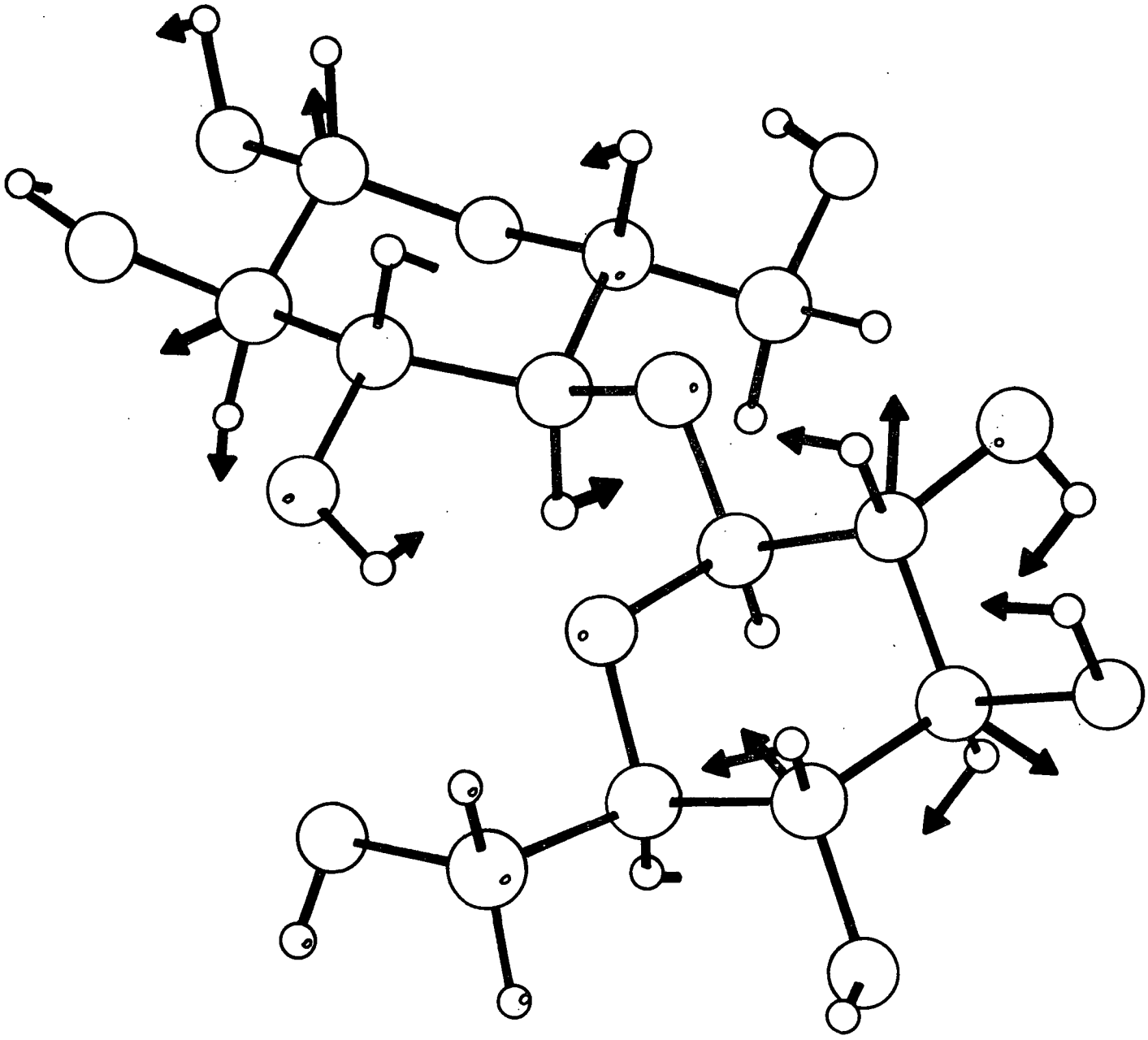
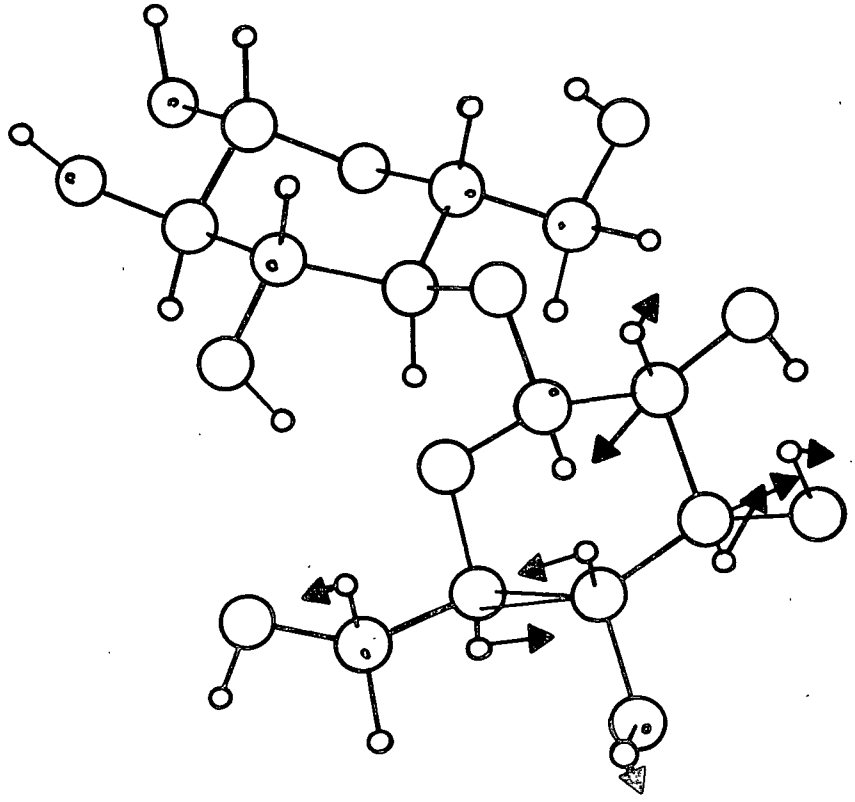


Figure 28. Displacements for the cellobiose calculated frequency at 1072 cm^{-1}.



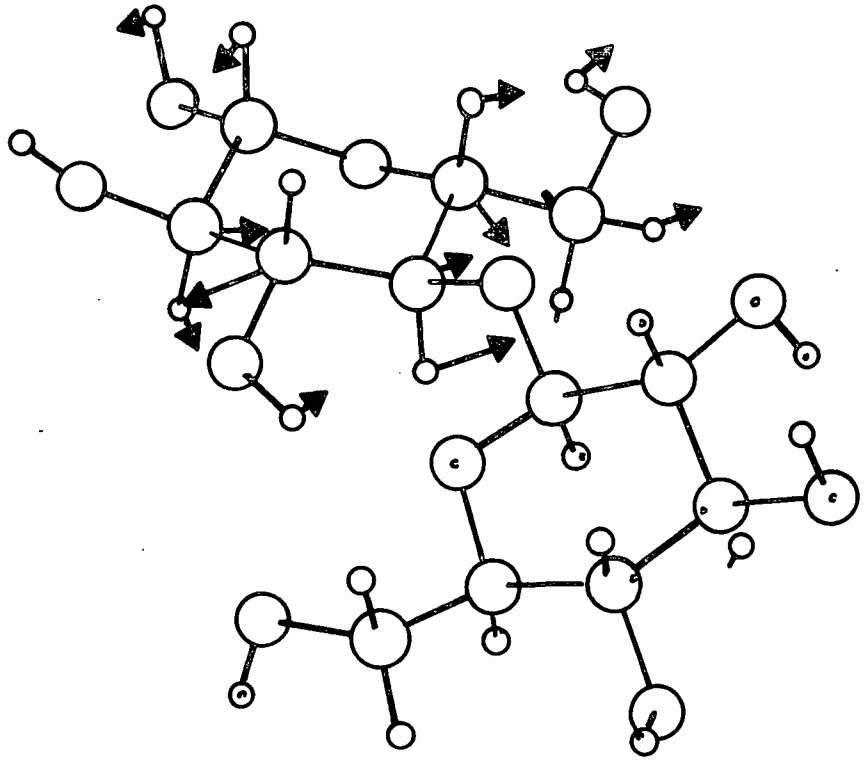


Figure 29. Displacements for a pair of complementary cellobiose calculated frequencies at 1110 and 1107 cm^{-1}.

Further evidence of this high degree of coupling in this region can be obtained from the sums of the PEDs. A measure of the degree of coupling for a single calculated frequency may be obtained from the deviation of the sum of the PED from 100% for that frequency. Figure 30 illustrates the sum of the PEDs plotted against wavenumber for the calculated frequencies. The PED is determined solely from the diagonal force constants and the more the sum of the PED deviates from 100%, the greater the amount of coupling for that calculated frequency. The sums of the PED in the region 1100-1000 cm^{-1} deviate significantly from 100% with the greatest sum of PED exceeding 150%. Although no absolute degree of coupling can be assigned, this indicates a high degree of coupling between the motions composing the normal modes in this region, namely C-C stretching, C-O stretching, and heavy atom bending.

It is evident from the data that both the C-O and C-C stretching motions are contributing to the calculated frequencies in this region and that they are highly coupled. Both of these factors contribute to the intense bands observed in this region for both the infrared and the Raman spectra of the disaccharides. The C-O stretching motions are responsible for the intense bands in the infrared due to the large change in dipole moment for these motions. The C-C stretching motions are responsible for the intense bands in the Raman due to a large change in polarizability associated with these motions.

The frequencies calculated in the region 890-850 cm^{-1} are dominated by the methine bending motions localized at the hydroxymethylene positions in all the disaccharides except xylobiose where this group is not present. In xylobiose, the methine bending motions at the C5 positions dominate this region. The displacements associated with the cellobiose normal mode calculated at 869 cm^{-1} are illustrated in Fig. 31. The displacements can be seen to be localized at the

C6 positions with minor motions throughout the rest of the molecule. The implications of this observation relative to the spectra of higher molecular weight compounds will be discussed following the discussion of the effect of conformational changes on the calculated frequencies for cellobiose.

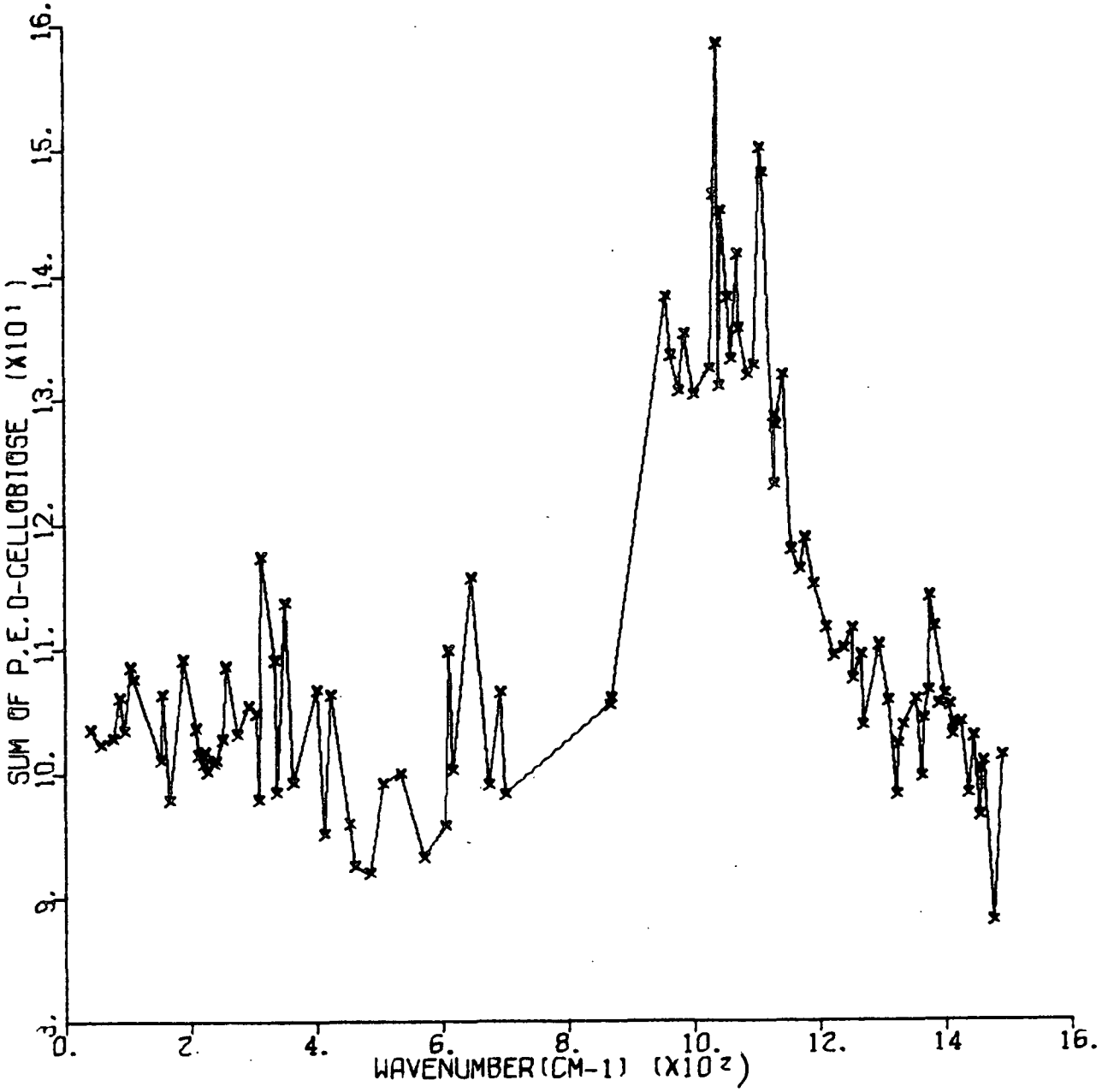


Figure 30. Sum of potential energy distributions versus wavenumber.

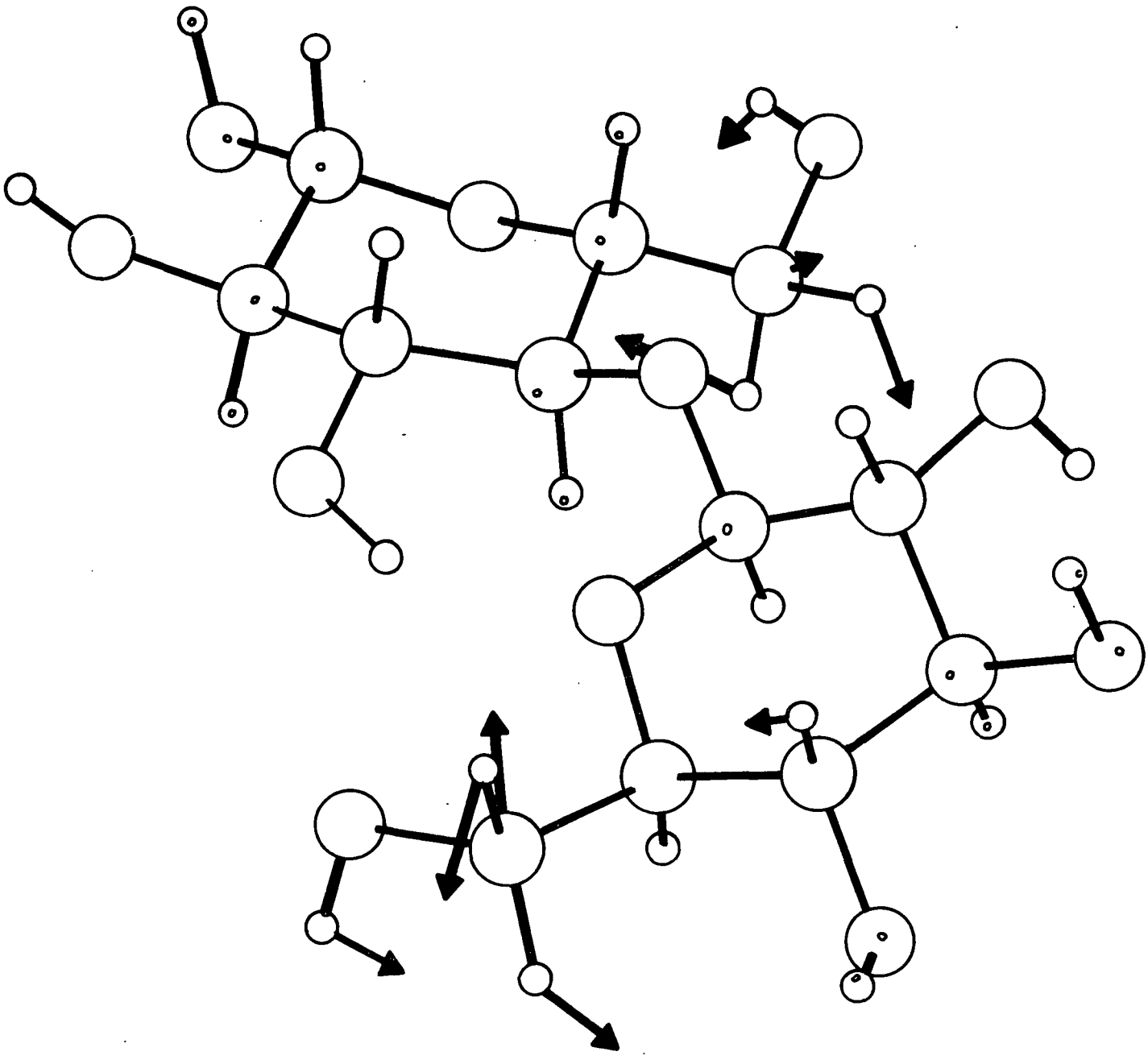


Figure 31. Displacements for the cellobiose calculated frequency at 869 cm^{-1}.

Region IV: 850-400 cm^{-1}

The observed frequencies for this region are relatively well spaced and of moderate intensity in both the infrared and Raman spectra of all the disaccharides. The calculated frequencies for this region are also well spaced and are a good representation of the general distribution of the observed bands. The potential energy distributions show this region to be dominated by heavy atom (C-C-O, O-C-O, and C-O-C) bending motions and C-O and C-C stretching motions. The relative contributions are illustrated in Fig. 32. The potential energy distributions are very delocalized for this region indicating very complex molecular motion for the normal modes in this region. A representative displacement drawing for this region is presented in Fig. 33.

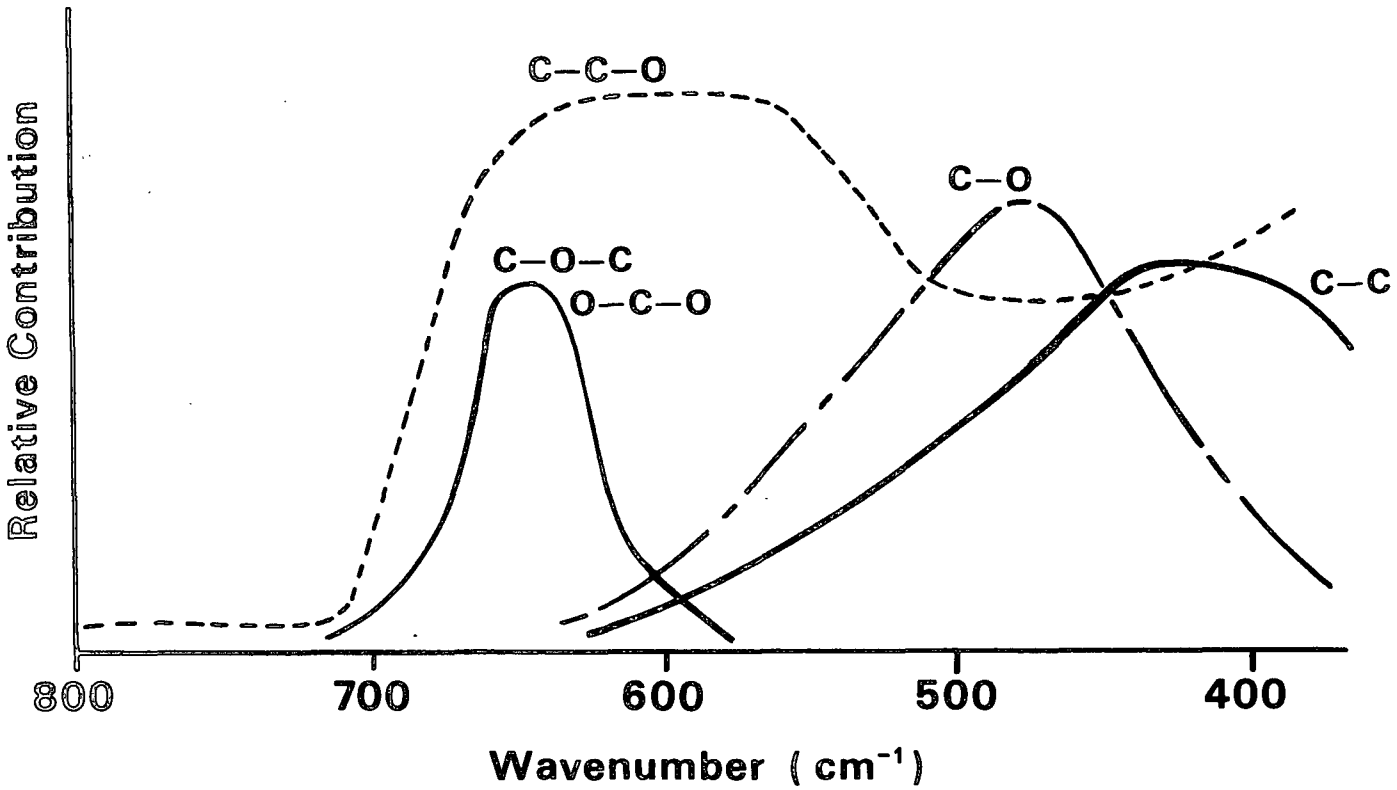


Figure 32. Relative potential energy contributions - Region IV.

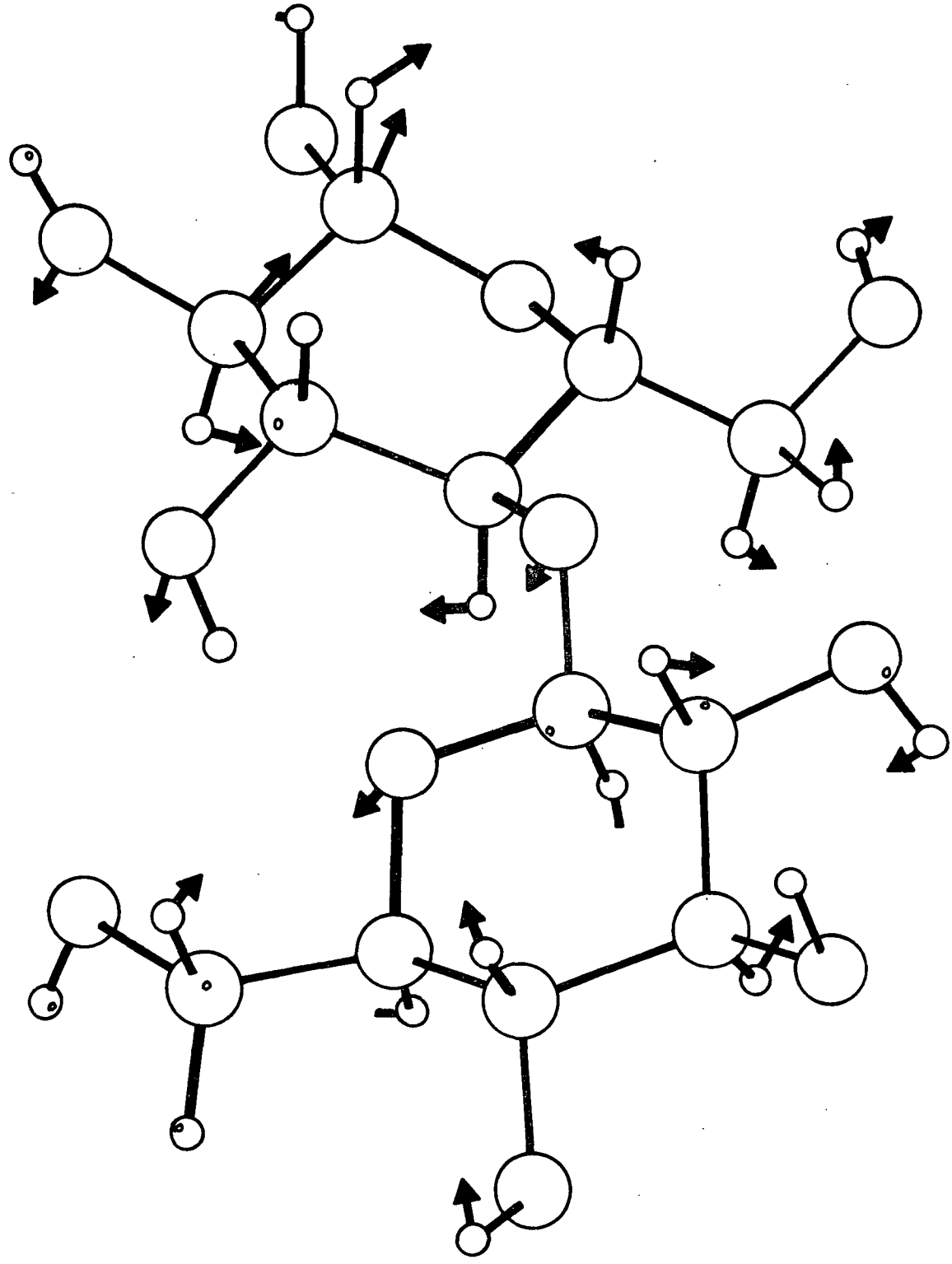


Figure 33. Representative normal mode displacements for Region IV - cellobiose normal mode at 460 cm^{-1}.

Also included in the infrared spectra of the disaccharides for region IV are bands corresponding to hydroxyl out of plane bending motions at approximately 750 cm^{-1}. These bands do not have corresponding Raman bands and are very sensitive to temperature, shifting as much as 50 cm^{-1} upon cooling to liquid nitrogen temperature (approximately -180°C). This motion was not reproduced by the normal coordinate calculations for the disaccharides nor was it reproduced for the hexoses by Wells' calculations.[24] This inability of the NCA to reproduce this observed frequency has been criticized by Michell.[40] The high temperature sensitivity of these bands indicates that they are most likely highly influenced by the intermolecular hydrogen bonding in the solid or other crystal affects. These types of interaction were not included in the normal coordinate calculations due to the assumption of a totally isolated molecule. Therefore, it is not surprising that this type of motion is not reproduced by the calculations.

Region V: 400-300 cm^{-1}

Figure 34 illustrates the major components for the motions calculated for the vibrations contained in Region V. The motions involved for the frequencies in this region are primarily C-C-C and C-C-O bending motions. These motions are customarily referred to as either ring breathing or ring bending motions. The observed frequencies for this region are of moderate to strong intensity in the Raman and of moderate intensity in the infrared spectra of all the disaccharides. As in Region IV, the motions for this region are very complex involving almost every heavy atom in the molecule. The distinguishing feature of the motions for the normal modes in this region is that there is significant motion of the oxygen atom at the glycosidic linkage. The displacements for the cellobiose normal mode calculated at 361 cm^{-1} are illustrated in Fig. 35 and demonstrate the significant motion of the glycosidic oxygen. Since the frequencies in this region are represented by motions of the glycosidic oxygen, any

change in either the geometry or hybridization of the glycosidic linkage should have an effect on the position and intensity of observed frequencies between 400 and 300 cm^{-1}. The reader is reminded of the same type of sensitivity in Region III (1150-800 cm^{-1}) but for a different reason, namely the coupling of motions through the linkage.

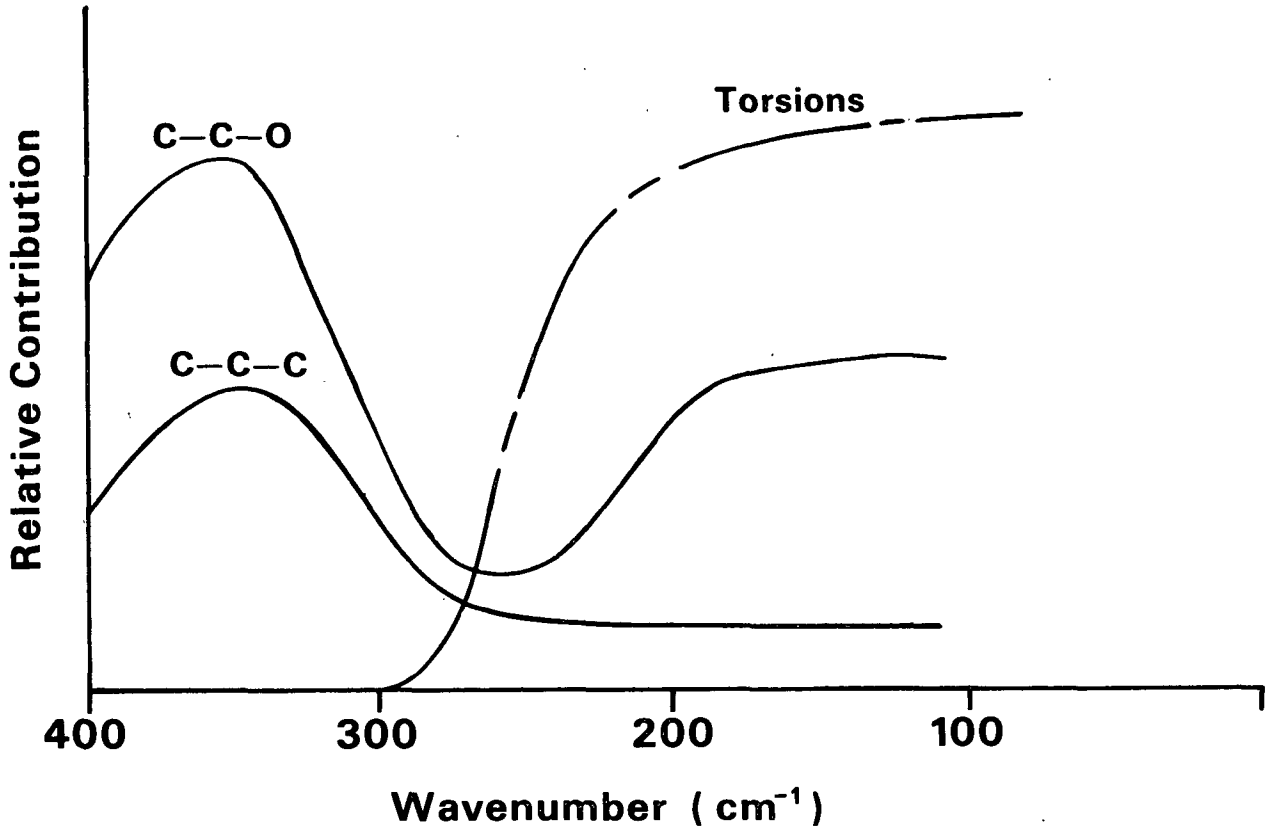


Figure 34. Relative potential energy contributions for Regions V and VI.

Region VI: Below 300 cm^{-1}

The major components calculated for the molecular motions for this region are the torsional motions with some contributions from the C-C-O bending motions. The relative contributions for this region are illustrated in Fig. 34. The torsions consist of out of plane bending motions about the C-C and C-O bonds within the molecules. The observed frequencies for this region are weak in both the infrared and Raman with only a few exceptions.

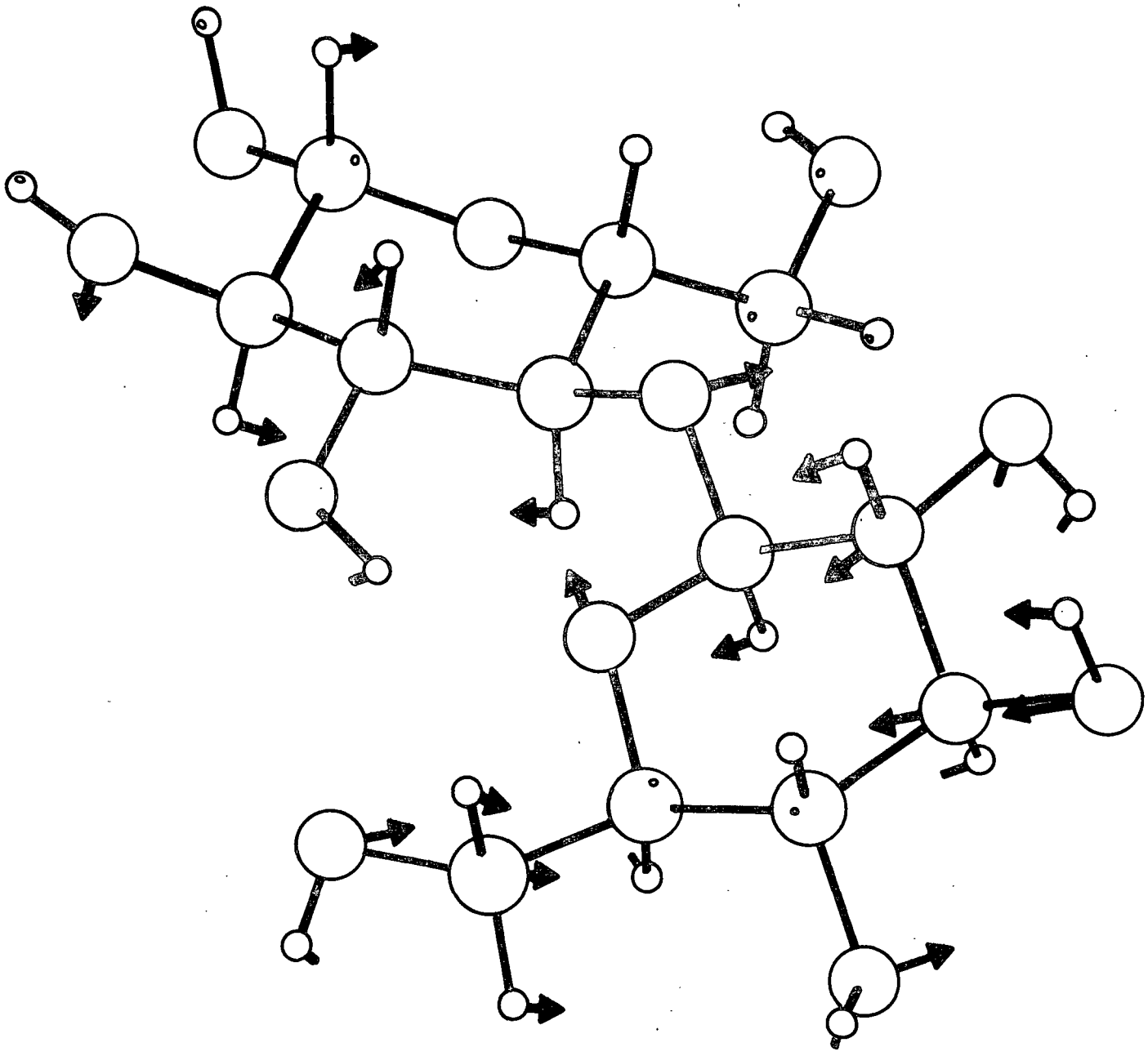


Figure 35. Displacements for the cellobiose calculated frequency at 361 cm^{-1}.

Also included in this region are long range couplings and crystal vibrations which cannot be reproduced by NCA. These bands can be visualized as standing waves within a crystal lattice. The assumption of isolated molecules does not allow the consideration of these effects and thus their nature cannot be interpreted based on the NCA results.

EFFECT OF VARYING THE CONFORMATION ON THE CALCULATED FREQUENCIES OF CELLOBIOSE

Introduction

The differences between the structures of cellulose I and cellulose II has been the subject of much debate. Some authors attribute the differences to the packing arrangements of the chains, either parallel or antiparallel.[41-44] Others contend that the differences lie in the conformations of the pendant hydroxymethylene units along the chains.[45,46] The structural differences between cellulose I and cellulose II have also been attributed to differences in the dihedral angles about the bonds at the glycosidic linkages along the chains.[47-51] These interpretations have been based primarily on crystallographic data and the ability of the models to simulate observed changes between the polymorphs. Another observed difference between cellulose I and cellulose II lies in their vibrational spectra.[47] The reader is referred to reference 48 for a discussion of these differences. These various hypotheses may be evaluated by the determination of the effect of the proposed changes on the calculated frequencies for a repeating unit of cellulose, namely cellobiose.

It is not possible to determine the effect of parallel versus antiparallel chain packing due to the assumption of an isolated molecule in the normal coordinate calculations. The ability of the NCA calculations to reproduce the spectra of the disaccharides using the isolated molecule assumption indicates

that the crystal effects have only minor impact on the actual frequencies. Therefore, if only the packing arrangement, and not the conformation of the chains, were altered, only slight changes in the vibrational spectra between cellulose I and cellulose II should be observed. Significant differences do occur between the spectra which cannot be explained solely by changes in packing arrangement. Other changes, namely conformational changes of the cellobiose model, were investigated in an attempt to explain the origin of the observed spectroscopic differences.

In order to determine the effects of conformational changes on the calculated frequencies of cellobiose, calculations were performed for cellobiose structures in which the C-O-C bridge angle, the hydroxymethylene conformations, and the dihedral angles about the bonds at the glycosidic linkage were systematically varied. These areas of conformational changes are illustrated in Fig. 36. The results of these sets of calculations will be treated separately.

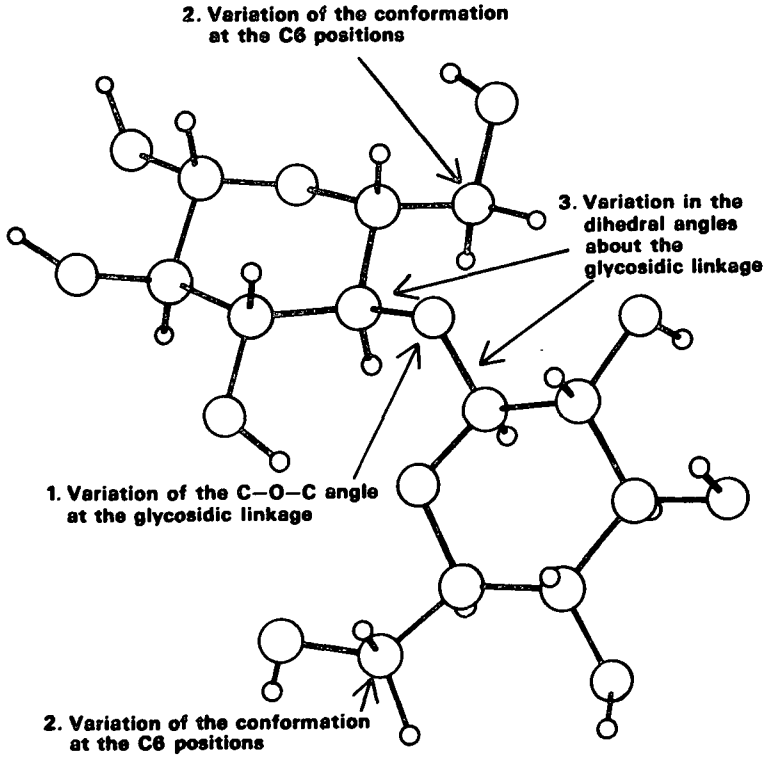


Figure 36. Areas of conformational changes in cellobiose.

Variation of the C-O-C Angle at the Glycosidic Linkage

The C-O-C angle at the glycosidic linkage in cellobiose, being 116° is apparently in a strained state as compared to a "free" equilibrium value of 113°. The free value is taken as that found in the ethers. This strained situation can be visualized as a result of the Van der Waals repulsive forces between the hydrogen atoms attached to the carbons on either side of the glycosidic linkage. A comparison of any changes in the calculated frequencies with the observed differences between the cellulose I and cellulose II spectra could suggest whether the nature of the cellulose polymorphy might lie in the C-O-C bridge angle at the glycosidic linkage. It was decided to perform a series of calculations in which the C-O-C angle was 115°, 117°, and 118° to determine if changes in the calculated frequencies might occur. The results of these calculations showed that there was very little change in the calculated frequencies when the C-O-C angle was changed. The possibility that changes in the C-O-C bridge angle might account for the nature of the cellulose polymorphy was pursued no further after these results.

Variation of the Hydroxymethylene Conformations

One of the differences between the observed vibrational spectra of cellulose I and cellulose II is the intensity of the band observed at 890 cm^{-1}. This band is weak in the infrared and very weak in the Raman spectra of cellulose I. It increases in intensity for cellulose II becoming a sharp band of moderate intensity in the Raman. These differences are shown in Fig. 37. The normal coordinate analysis for cellobiose shows a calculated frequency in this region which is comprised of primarily methine bending motions localized at the hydroxymethylene positions. It is likely that the motions associated with this band in both cellulose I and cellulose II are comprised of similarly localized molecular

motions. Zhbankov[17] observed this effect in his normal coordinate calculation for cellobiose. He used this observation to confirm his contention that the differences between cellulose I and cellulose II lie in the conformations of the pendant hydroxymethylene units. However, Zhbankov did not present any confirming evidence in the form of calculations for alternative conformations of these hydroxymethylene units. A series of calculations for different hydroxymethylene conformations were performed in this study to test this hypothesis.

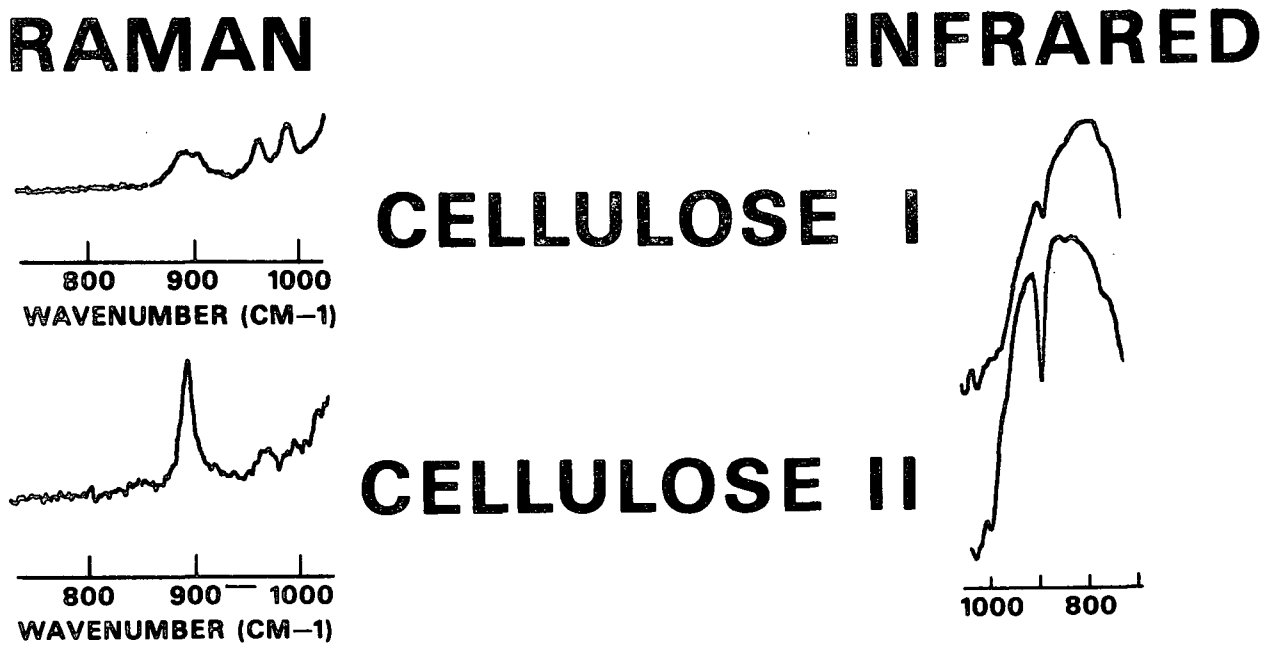


Figure 37. Infrared and Raman spectra for cellulose I and cellulose II - 800-900 cm^{-1}.

The three preferred hydroxymethylene conformations for glucose are shown in Fig. 38. The conformations are labeled according to the relative position, either gauche (g) or trans (t), of the C6-O6 bond to first the C5-O5 bond and then the C5-C4 bond. For example in the gt conformation, the C6-O6 bond is

gauche to the C5-O5 bond and trans to the C5-C4 bond. The conformations considered in the calculations were the gt and gg conformations. The tg conformation was not considered due to the steric hindrance of the hydroxyl oxygen at C4. The gt conformation is the conformation present in the crystal structure of cellobiose and gg was considered an alternative conformation. This allowed four combinations of hydroxymethylene conformations for cellobiose which are illustrated in Fig. 39.

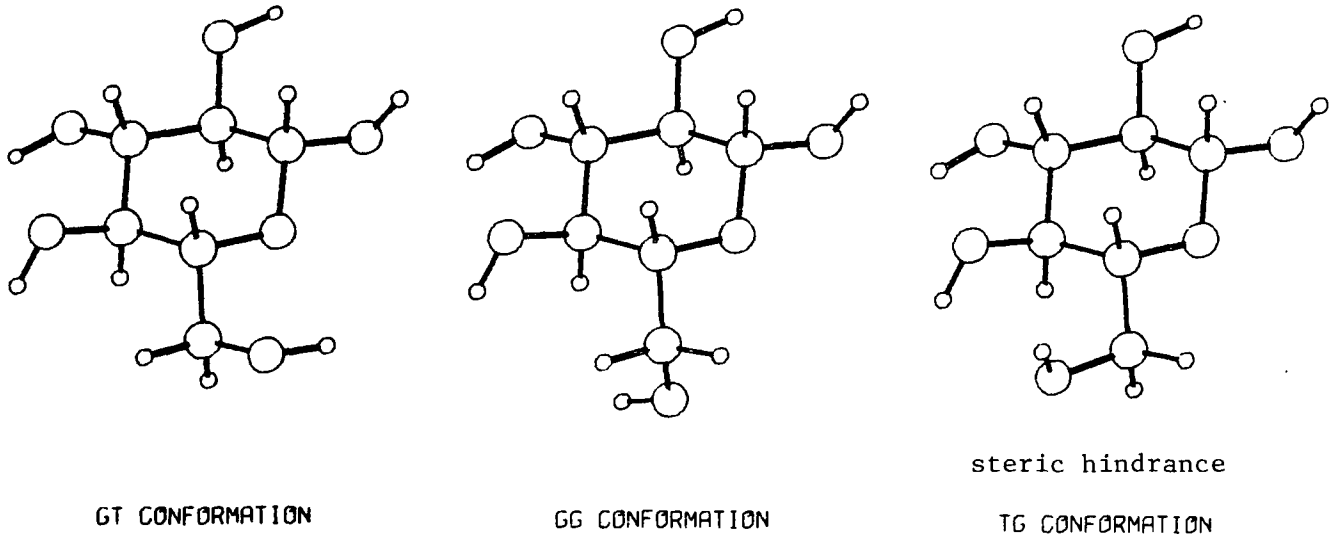
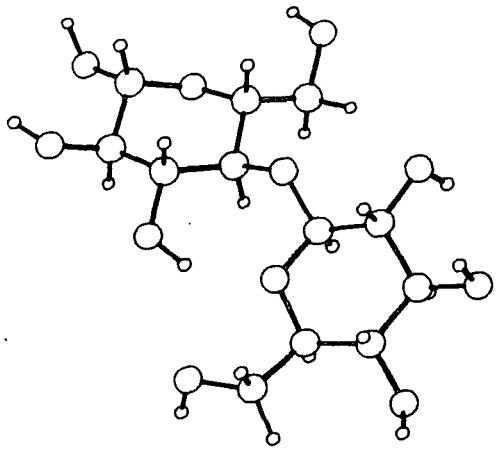
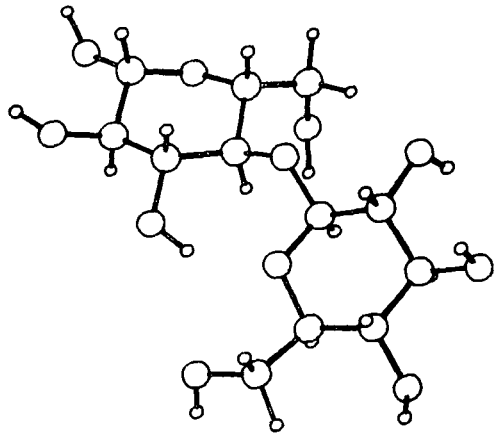


Figure 38. Hydroxymethylene conformations for glucose.

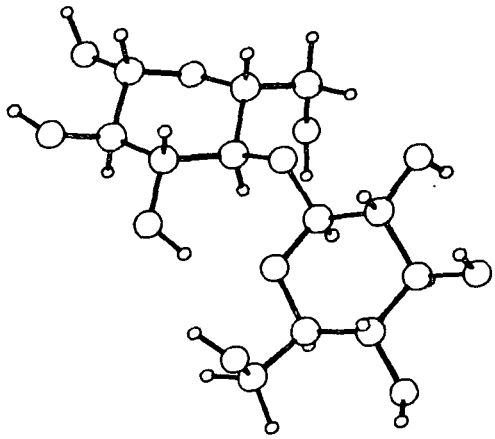
There were several changes in the calculated frequencies when the hydroxymethylene conformations were changed. The differences in the calculated frequencies for the four structures with varying hydroxymethylene conformation are shown in Fig. 40 along with a representation of the Raman spectrum of cellobiose. There were changes in the calculated frequencies in the low frequency region, 400-300 cm^{-1} which appeared to reflect the differences between the observed frequencies for cellulose I and cellulose II. However, the largest differences between the calculated frequencies for the structures occurred in the region 900-650 cm^{-1}. These changes in the calculated frequencies resulted



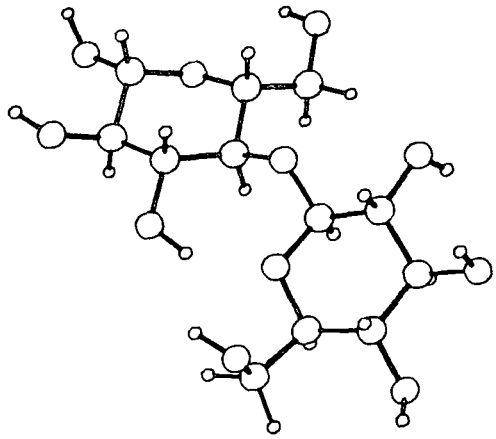
CELLOBIOSE STRUCTURE A - CRYSTAL STRUCTURE



CELLOBIOSE STRUCTURE B - C6' GG



CELLOBIOSE STRUCTURE C - BOTH C6'S GG



CELLOBIOSE STRUCTURE D - C6 GG

Figure 39. Four cellobiose structures with various hydroxymethylene conformations.

in the appearance of frequencies where there were no corresponding observed fundamental frequencies in cellobiose, cellulose I or cellulose II. This shift of frequencies into a spectral region void of observed bands makes the possibility of hydroxymethylene rotation suspect as a cause for the difference in celluloses I and II. The composition of the shifted frequencies between 420 and 350 cm^{-1}

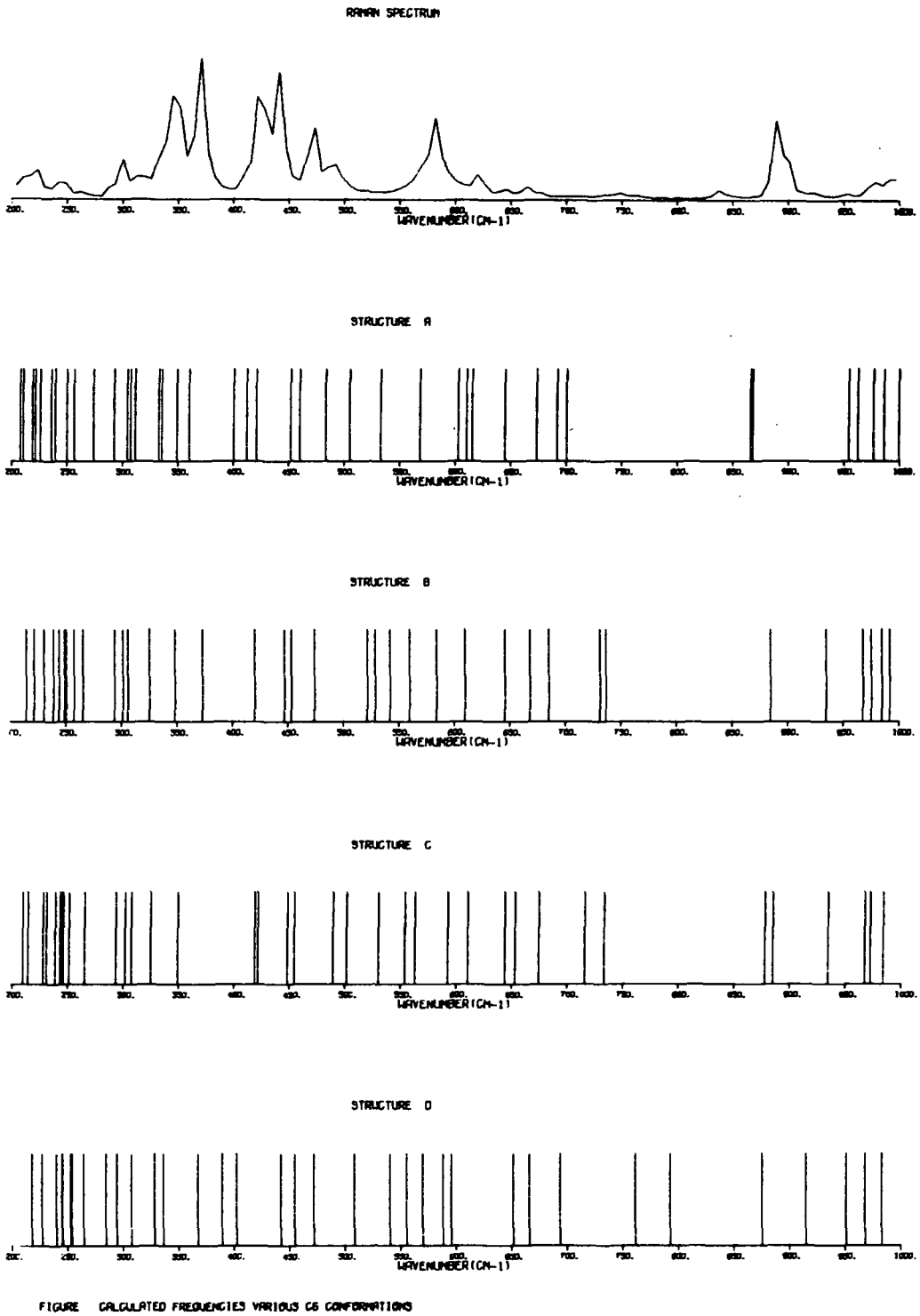


Figure 40. Comparison of calculated frequencies for cellobiose structures having different hydroxymethylene conformations.

further disputes this possibility. Table 5 presents the PED's for structure B between 420 cm^{-1} and 350 cm^{-1}. It can be seen that the calculated normal mode at 372 cm^{-1} is highly torsional in nature. This is a very high frequency to be primarily torsional and raises the question of either an unrealistic structure or an error in the calculations. Several checks of the calculations resulted in serious questions on the merit of the structure being representative. On the basis of these observations it was concluded that the differences between the vibrational spectra of cellulose I and cellulose II cannot be accounted for by changes in the conformations at the hydroxymethylene units.

Table 5. PED's for structure B - 420 cm^{-1} to 350 cm^{-1}.

Frequency = 420 cm^{-1}

18.08	C5-C5-O5
16.98	C1-C2-C3
7.64	O1-C1-C2
7.61	O5-C1-C2
7.34	C5-C6-O6
4.82	C2-C3-C4
4.07	O2-C2-C3
2.75	C2'-C3'
2.33	C2-C3-O3
2.25	C1O4'-C4'
2.23	C1'-O1'
1.81	C1-C2-O2

The sum of the PES is 99.13

Frequency = 372.80 cm^{-1}

32.64	TO4'-C1
17.10	TC4'-O4'
6.41	TC2'-C3'
4.29	O5C21O4'
3.52	H1C21O4'
3.07	TC4-C5
3.05	O2C22-C3
2.89	C26C5-O5
2.41	C23C4-O4
2.34	TO5-C1

The sum of the PES is 104.01

Frequency = 347.80 cm^{-1}

16.99	C1-C2-O2
11.74	C2-C3-O3
11.72	C3-C4-O4
2.16	O5-C1-O1
1.99	C5-O5
1.82	C5-C6

The sum of the PES is 99.18

Frequency = 350.10 cm^{-1}

13.00	C21C2-C3
10.10	O4C24-C5
9.77	C2-C3-O3
8.37	C3-C4-O4
6.63	O3C23-C4
6.20	C1-C2-O2
3.56	O4'C1-O2
3.04	C22C3-O3
2.82	H3-C3-C4
2.59	O5-C1-C2
2.57	C24C5-H5
2.37	C4-C5
2.09	O2C22-C3
1.99	C3-C4

The sum of the PES is 113.71

An interpretation of the difference in intensity of the observed band at 890 cm^{-1} in cellulose I and cellulose II will be presented later. The differences are interpreted in terms of changes in the dihedral angles about the bonds at the glycosidic linkage. A basis for discussing these changes about the glycosidic linkages must first be developed.

Effect of Changes in the Conformation About the Glycosidic Linkage Upon the Calculated Frequencies for Cellobiose

A proposed difference in the structures of cellulose I and cellulose II is a difference in the dihedral angles about the bonds participating in the glycosidic linkages.[47-51] The potential energy of interatomic interaction, or conformational energy, has been calculated for various values of the dihedral angles about the glycosidic linkage in cellobiose by Reese and Skerrett.[52] Their work forms a base for the discussion of changes in the dihedral angles about the bonds at the glycosidic linkage. Their results have been used extensively in this work. Their investigation will be discussed before proceeding to the effect of conformational changes at the glycosidic linkage upon the calculated frequencies of cellobiose.

The work on the conformational energy of cellobiose[52] was performed in three stages. First, a definition of the zero values and positive sense of rotation for the dihedral angles, ϕ and ψ, was adopted. This convention is illustrated in Fig. 41. The second stage was the determination of the conformational energy at 10° intervals of ϕ and ψ. The conformational energy was obtained by summing the potential energy of interaction for all the atoms in cellobiose. The potential energy is determined by the spatial separation of the atoms. Various functions for the interaction term were used[52] yielding comparable results. These calculations resulted in the elimination of 96% of the possible conformations on

the basis of hard sphere overlap. This is illustrated in Fig. 42 with the allowed area outlined by the solid line and a marginally allowed region outlined by the dotted line. Finally, calculations were performed within the allowed region at 2° intervals of both ϕ and ψ, and contour maps were constructed from the sums of the potential energies. Figure 43 is a section of the allowed region with the contour lines of conformational energy shown. It should be noted that there are two energy minima in the contour map. The importance of this will be emphasized later in the discussion.

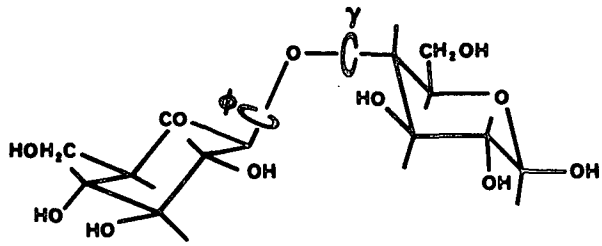


Figure 41. Scheme for variation of bridge-angle geometry in cellobiose $\phi = \psi = o°$. Rotations are positive when each residue is rotated in the direction shown about its bond to the bridge oxygen.

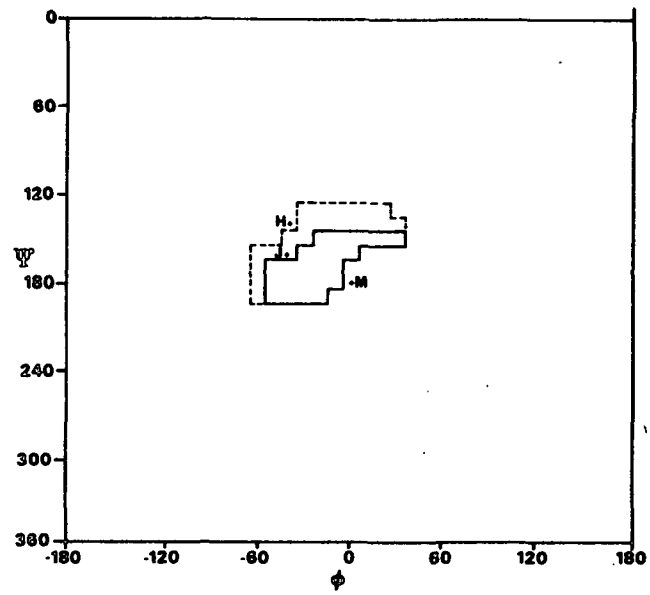


Figure 42. Allowed (enclosed by solid line) and marginally allowed (broken line) cellulose conformations, using Set 4 coordinates with all relevant contacts considered. Cellobiose (J), cellulose (H), and cellulose (M) are marked as J, H, and M, respectively.

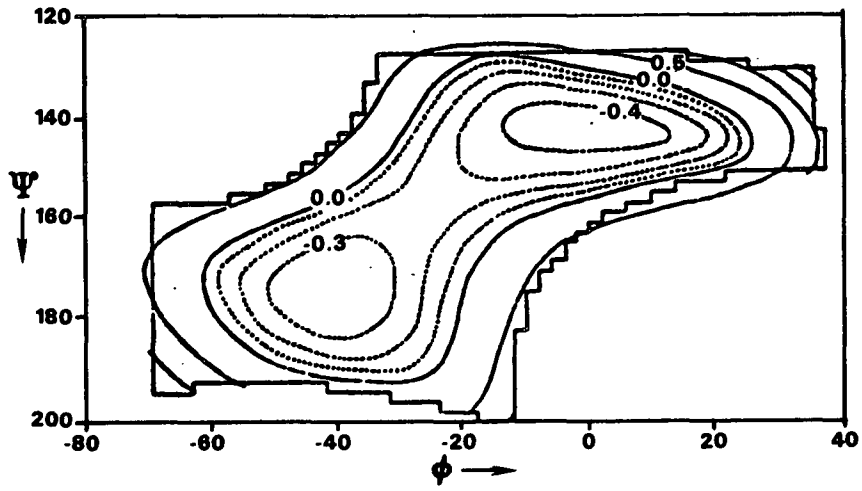


Figure 43. Conformational energy map (from Reese and Skerrett[53]). The units of energy on this map are Kcal/mole.

Figure 44 shows a section of the allowed region containing further structural information. The two energy minima are illustrated by the solid curve. The broken lines in Fig. 44 define the conformations with the same projected repeat distance and the solid diagonal lines define those conformations with the same screw symmetry. These values are important for comparison with limited crystallographic data. The O3'...O5 distances for the limits for the existence of an intramolecular hydrogen bond are denoted by the dotted curves. It should be noted that the two conformational energy minima are approximately equally spaced from the line defining the structures with a twofold screw symmetry. The representation of the crystal structure of cellobiose (marked J) is near one of the minima and the representation of the crystal structure of methyl β-cellobioside (marked W) is near the other.

The five structures used in the calculations for cellobiose are represented on Fig. 44 by the letters A through E. All of these structures lie within the region bordered by bold lines and labeled "area of interest." This region is defined by the lines defining the structures having projected repeat distances of 5.1 and 5.2 Å and by the lines defining the O3'...O5 distances of 2.5 and 2.8 Å. The repeat distances are consistent with the accepted value of 10.3 Å for

cellulose obtained from the crystallographic data. The O3'...O5 distances of 2.5 and 2.8 Å are reasonable limits for the existence of an intramolecular hydrogen bond between the rings as shown in Fig. 45. The effect of this intramolecular hydrogen bond was not included in the potential energy calculations by Reese and Skerrett. It if had been included, it would have reduced the total conformational energy for the region which allows the hydrogen bond. This reduction in total conformational energy would cause the two energy minima to expand toward, and probably into, the region marked "area of interest" in Fig. 44. This together with the agreement with the crystallographic data, implies that the area on the ϕ-ψ contour map considered in these calculations represents the most likely conformations about the glycosidic linkages for cellulose (both cellulose I and cellulose II).

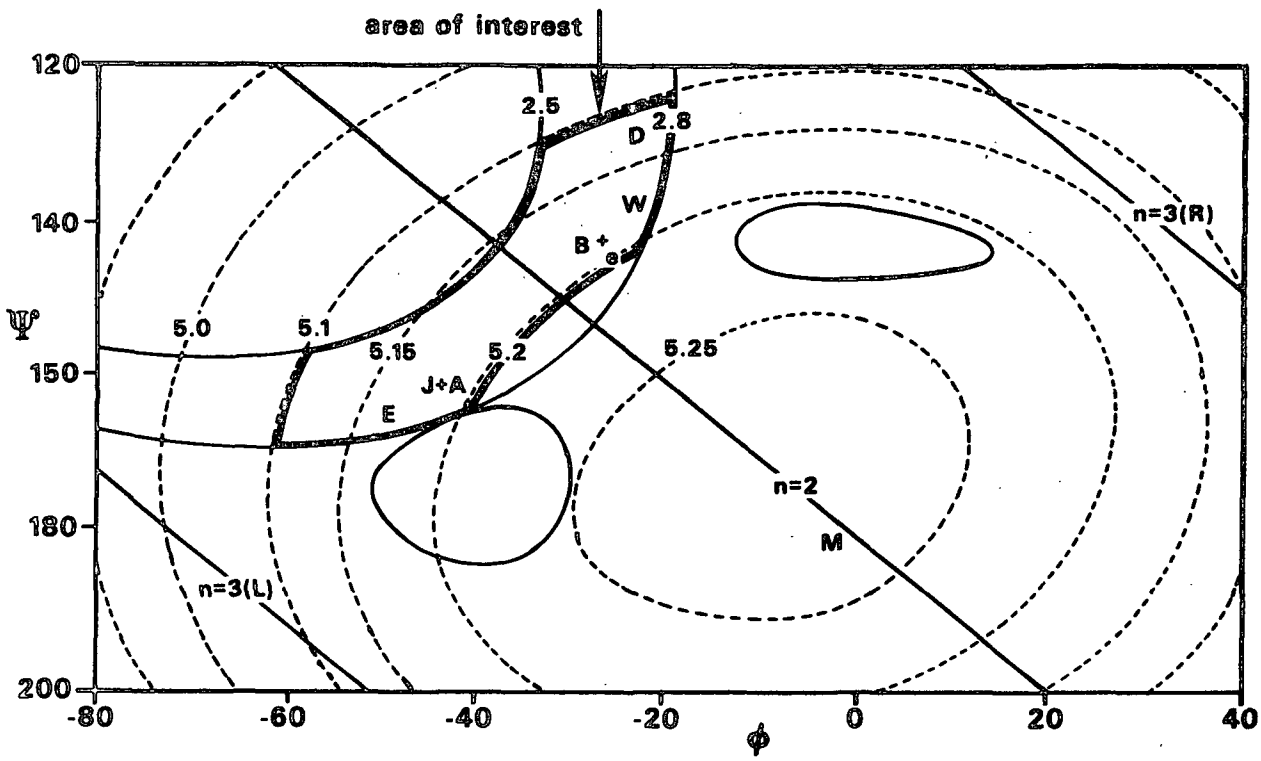
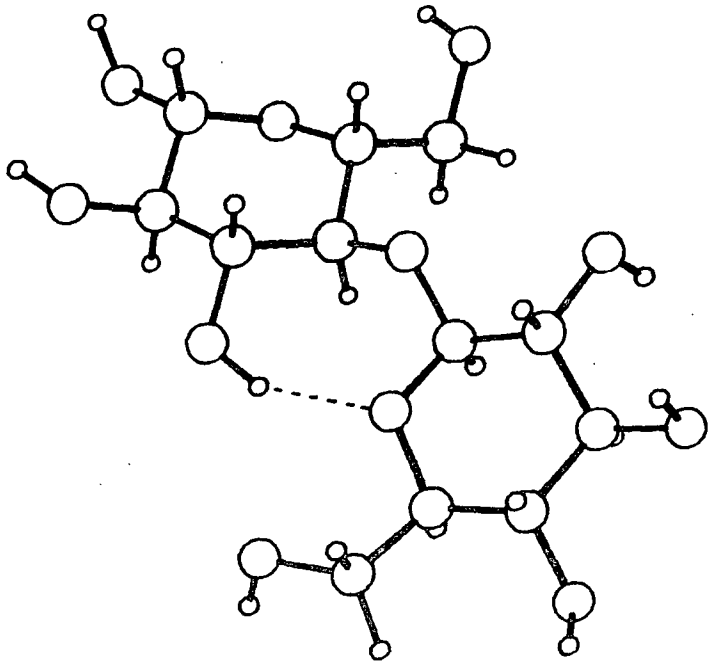


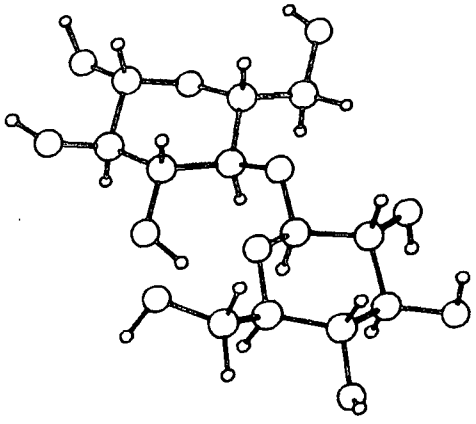
Figure 44. ϕ-ψ Contour map containing further structural information.



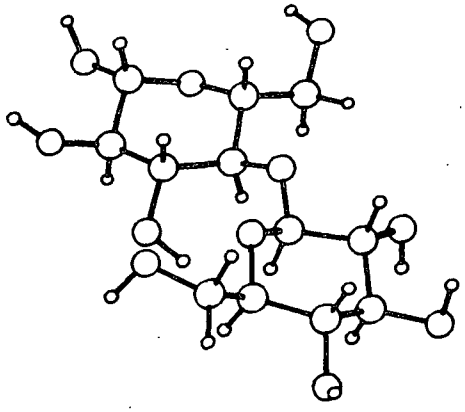
CELLOBIOSE CRYSTAL STRUCTURE

Figure 45. Cellobiose structure with intramolecular hydrogen bond illustrated.

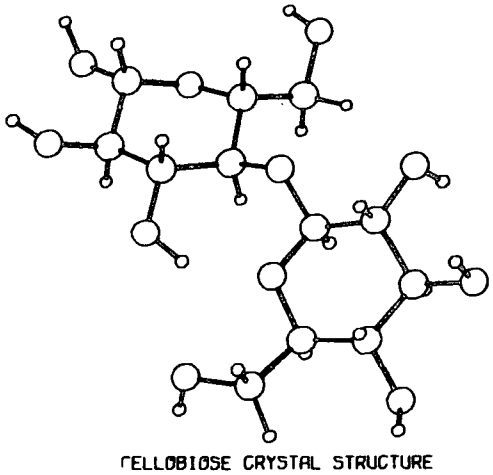
The effect of conformational changes about the glycosidic linkage upon the calculated frequencies for cellobiose was localized in the frequency range 400-300 cm^{-1}. This is the same region in which differences were observed between cellulose I and cellulose II. The five structures used in the calculations are shown in Fig. 46. Again, their representations on the ϕ-ψ contour map are labeled A through E on Fig. 44. It can be seen that the structures represent both sides of the twofold screen symmetry line. The results of the NCA calculations for these structures reflect the observed differences between the observed spectra of cellulose I and cellulose II in the region 400-300 cm^{-1}. The structures having a right-handed screen symmetry (like methyl β-cellobioside) have calculated frequencies similar to those of cellulose I, and the structures with left-handed screw symmetry (like cellobiose) have calculated frequencies similar to the observed frequencies of cellulose II. The calculated frequencies below



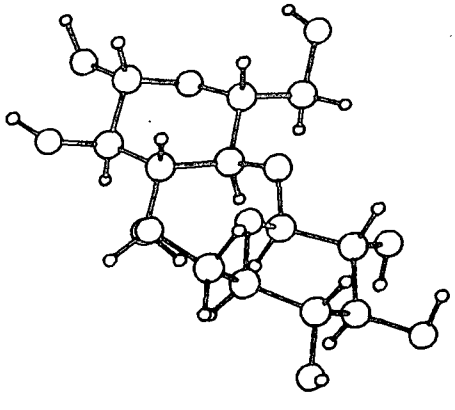
CELLOBIOSE STRUCTURE E



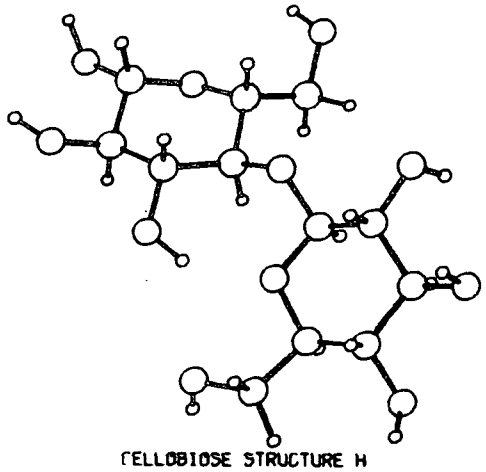
CELLOBIOSE STRUCTURE F



CELLOBIOSE CRYSTAL STRUCTURE



CELLOBIOSE STRUCTURE G



CELLOBIOSE STRUCTURE H

Figure 46. Five structures used in determining the effect of conformational changes about the glycosidic linkage on the calculated frequencies of cellobiose.

1000 cm^{-1} for the crystal structure of cellobiose and one cellobiose structure with a right-handed screw axis are presented in Fig. 47 along with a representation of the Raman spectrum of cellobiose. The representation is a computerized drawing of the Raman spectrum with the data digitized at 6.4 cm^{-1} intervals.

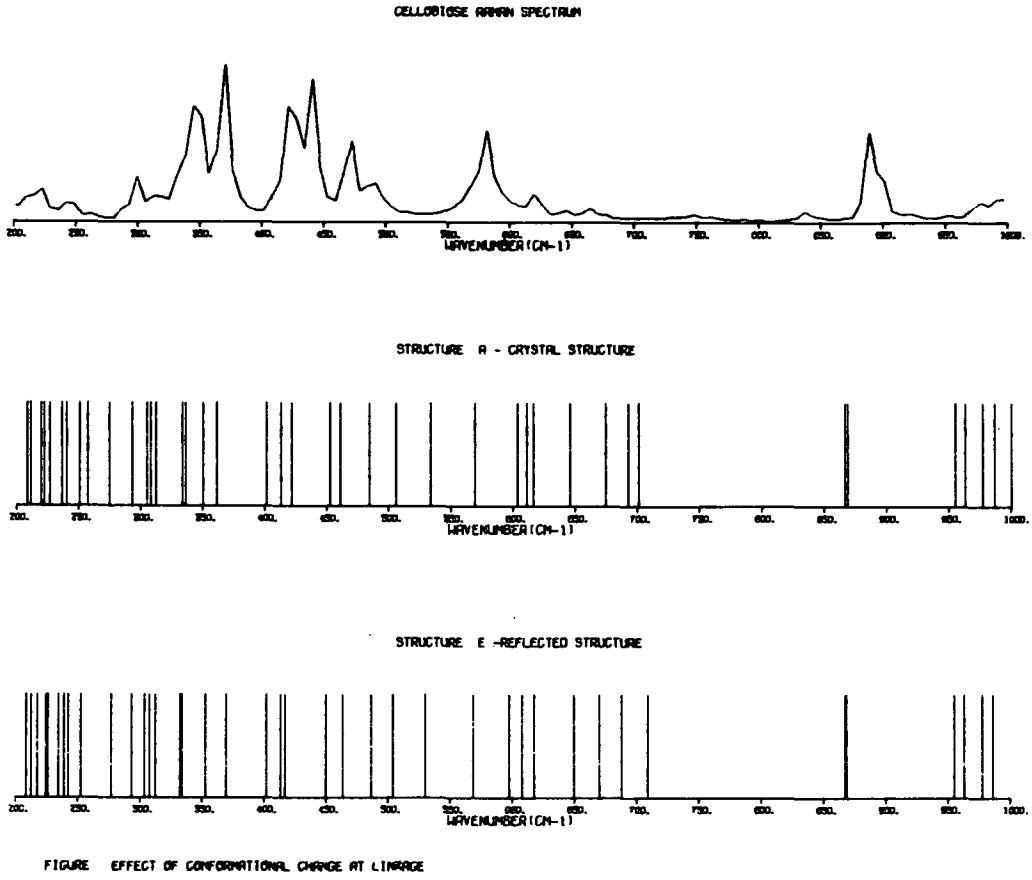


Figure 47. Comparison of calculated frequencies for cellobiose crystal structure and cellobiose structure with right hand screw symmetry with representation of cellobiose Raman spectrum.

The results of these calculations indicate that the differences between the observed vibrational spectra of cellulose I and cellulose II can be attributed to differences in conformation about the glycosidic linkages.

It is appropriate to comment on the difference in the intensity of the bands observed at 890 cm^{-1} for cellulose I and cellulose II. The normal coordinate analysis indicates that, for the disaccharides, bands in this region are composed of methine bending motions localized at the hydroxymethylene positions. This appears to be in conflict with the statement that differences in the observed spectra of cellulose I and cellulose II can be attributed to conformational differences about the glycosidic linkages. The conclusion, based on the normal coordinate calculations for cellobiose, was that cellulose I has conformation(s) about its glycosidic linkages similar to that found in methyl β-cellobioside, and cellulose II has conformation(s) similar to cellobiose. The differences in intensity of the observed band at 890 cm^{-1} may be interpreted in terms of conformational changes at the glycosidic linkages also. The interpretation is based on the type of intramolecular hydrogen bond that would be expected for the two types of conformations. The structures of methyl β-cellobioside and cellobiose are shown in Fig. 48 with the intramolecular hydrogen bonds indicated. The intramolecular hydrogen bond in methyl β-cellobioside is bifurcated, including the hydroxymethylene oxygen, whereas the cellobiose intramolecular hydrogen bond is isolated. The participation in the intramolecular hydrogen bond by the oxygen of the hydroxymethylene unit may account for the difference in intensity of the band at 890 cm^{-1}. In cellulose II, the hydroxymethylene units would be affected only by the intermolecular hydrogen bonding and the sharp band which is observed. The intermolecular hydrogen bond effects on this band in cellulose II are only minimal as indicated by the insensitivity of this band to changes in temperature. A high temperature sensitivity is indicative of a large influence by hydrogen bonding. In cellulose I, the motions of the hydroxymethylene units could be hindered by the intramolecular hydrogen bond because the bifurcation includes the oxygen at C6. This could account for

the reduced intensity of the observed band at 890 cm^{-1} in cellulose I with respect to the band for cellulose II. The perturbation by the bifurcated intramolecular hydrogen bond may be viewed as either coupling the hydroxymethylene motions to the motions in the neighboring ring or changing the atomic orbitals and therefore changing the shifts in polarizability and dipole moment which account for the bands.

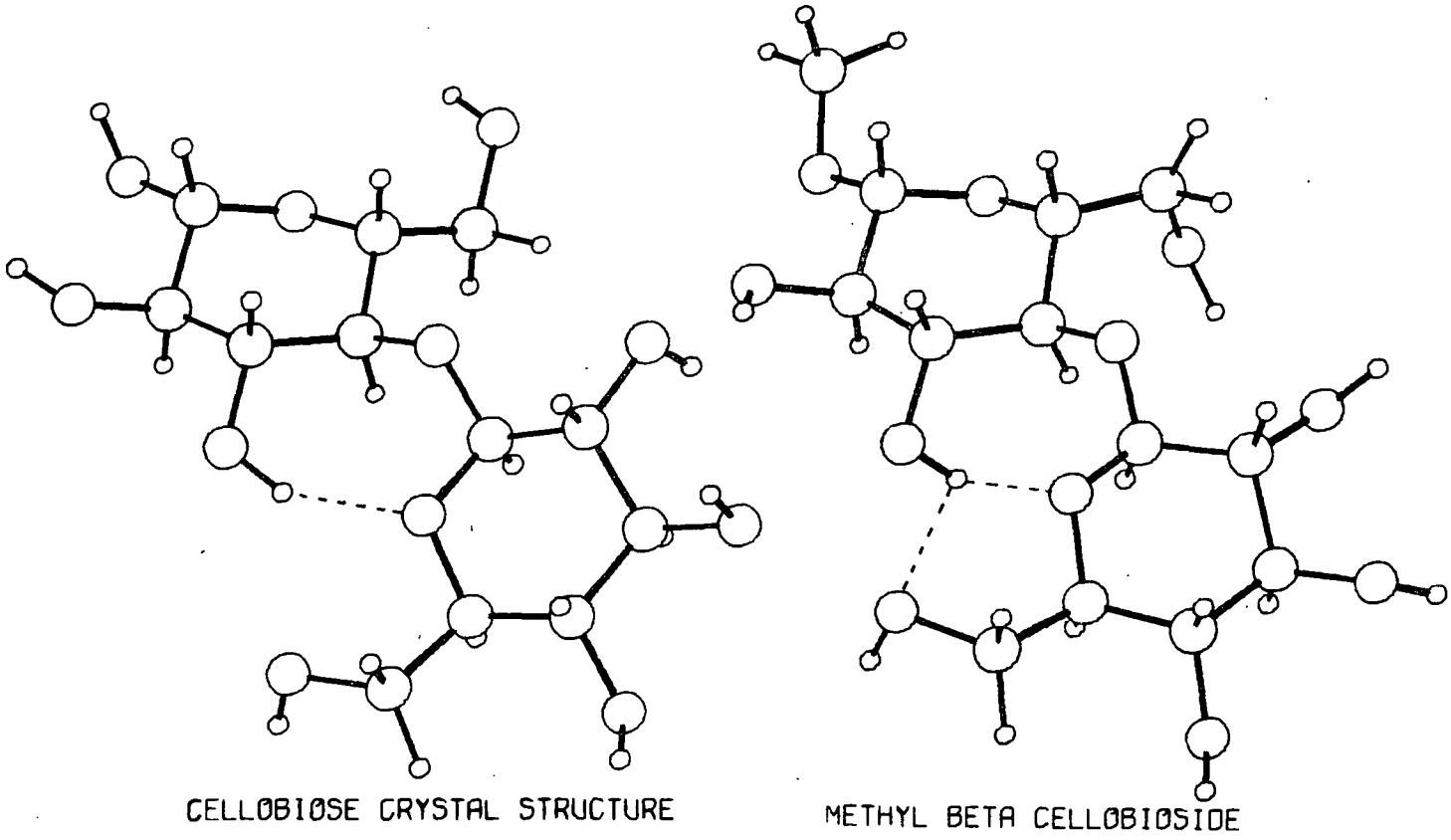


Figure 48. Structures of cellobiose and methyl β-cellobioside illustrating the intramolecular hydrogen bonds.

CONCLUSIONS

The results of this section of the thesis work show that normal coordinate analysis is a useful tool in the investigation of the vibrational spectra of molecules as complex as the disaccharides. The atomic motions associated with the vibrational frequencies below 1500 cm^{-1} are very complex with a high degree of coupling for the disaccharides. Calculations for a number of cellobiose conformations indicate that changes in the dihedral angles about the bonds at the glycosidic linkage result in changes in the calculated frequencies similar to the observed differences between the observed frequencies of cellulose I and cellulose II. Based on these calculations, it was concluded that the differences between the observed spectra of cellulose I and cellulose II can be attributed to changes in conformation about the glycosidic linkages. These differences in conformation about the glycosidic linkages can result in changes in the type of intramolecular hydrogen bond present in cellulose I and cellulose II. This difference in intramolecular hydrogen bonding can account for the observed difference in intensity of the bands observed at 890 cm^{-1} in cellulose I and cellulose II vibrational spectra.

SECTION III

INVESTIGATION OF THE VIBRATIONAL SPECTRA OF THE CELLODEXTRINS

INTRODUCTION

The results presented in Section II indicate that the differences in the vibrational spectra of cellulose I and cellulose II may be interpreted in terms of differences in conformation about the glycosidic linkages. The logical extension of these results is the investigation of the vibrational spectra of the higher cellodextrins with particular emphasis on the sensitivity of the calculated frequencies to changes in conformation about the glycosidic linkage. This has been done for cellotetraose with the results taken as representative of the behavior of the higher cellodextrins.

EXPERIMENTAL

Chromatographic Separation of the Cellodextrins

The method chosen for obtaining the cellodextrin samples was the chromatographic separation of hydrolyzed cellulose using a steric acid coated carbon/Celite column eluted with a water/ethanol mixture. This method has been described by Miller[53] and has been successfully used by others.[54] Recent advances in high performance liquid chromatography[55,56] appear to offer a favorable alternative to the method described below.

Column Preparation

Darco G-60 (400 g) and Celite 535 (400 g) were mixed with 3200 mL of 2.5% stearic acid in absolute ethanol for 30 minutes. This slurry was suction filtered on a 20 cm (8 inch) Buchner funnel and remixed with 3200 mL of 50% (v/v) ethanol in water saturated with stearic acid. The slurry was again

suction filtered on a 20 cm (8 inch) Buchner funnel and then placed in 2 liters of 10% (v/v) ethanol in water and allowed to settle overnight. The mixture was decanted with the thick slurry (adsorbent) set aside for use in packing the column. It is believed that a more uniform packing is obtained when the thick slurry is used as opposed to a dilute slurry.[57]

A glass chromatographic column, 150 cm x 5 cm inside diameter, with a medium porosity fritted glass plate, was used for the separation. The column was first washed with chromic acid, then nitric acid, and finally thoroughly washed with distilled water before being packed. Approximately 10 cm of Celite 535 were placed above the fritted glass plate in a slurry form to trap any carbon particles which might otherwise plug the fritted glass plate. The adsorbent was poured into the column and allowed to settle. The resulting supernatant liquid was forced through the column using air pressure (see Fig. 49) with care taken that the liquid level did not drop below the top of the adsorbent. The column was then washed with 4 liters of distilled water under 3 psi pressure and the flow rate was measured at 1.5 mL/min. The height of the adsorbent above the Celite plug was approximately 90 cm.

Hydrolysis of Cellulose

Whatman CF-1 cellulose powder (15 g) was first wetted with 100 mL of concentrated hydrochloric acid (37%, sp.gr. 1.19) in a 250 mL Erlenmeyer flask and stirred for one minute at room temperature to yield a uniform suspension. One hundred mL of fuming hydrochloric acid (42%, sp.gr. 1.21) was then added in one portion. The fuming hydrochloric acid was prepared by bubbling HCl gas from a tank into concentrated hydrochloric acid, cooled to 0°C in an ice bath, until the desired specific gravity was obtained. The resulting viscous solution was allowed to warm to room temperature and stand for a total of two hours. At that

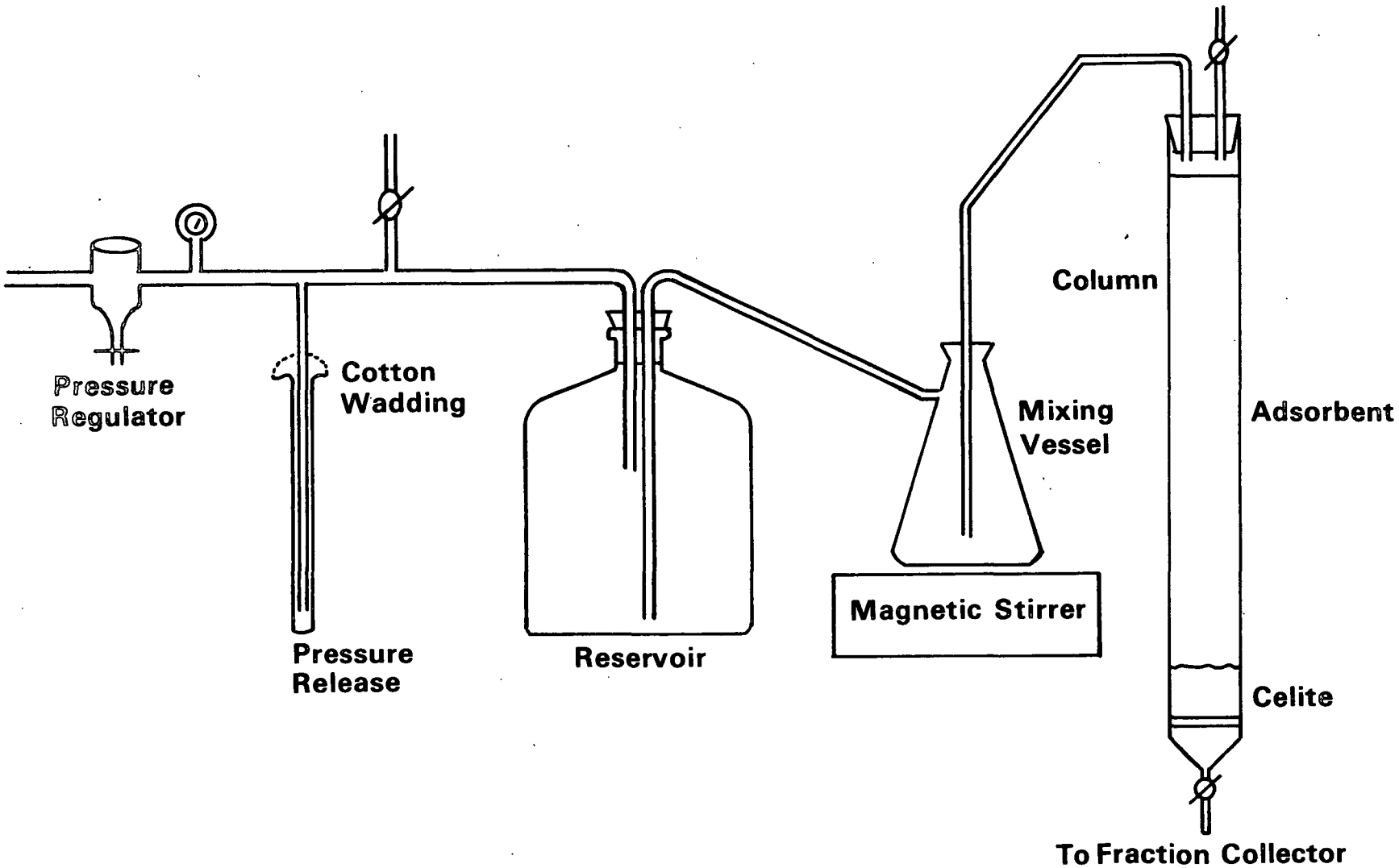


Figure 49. Diagram of separation apparatus.

time the solution was poured into 600 mL of ice-cold distilled water and neutralized to pH 5.5 with portion-wise addition of 210 g of sodium bicarbonate. The undesired gelatinous material was removed by suction filtration on Whatman No. 1 filter paper. The first portion was cloudy and returned to the filter. The filtration was very slow taking approximately 3 hours. The filtrate, or hydrolyzate, was used for the fractionation on the chromatographic column.

Fractionation of the Hydrolyzate

The fractionation apparatus is illustrated in Fig. 49 and the fractionation process was as follows. After the hydrolyzate was introduced into the column, the column was washed with 2.5 liters of distilled water to remove the salt and glucose present in the hydrolyzate. The top of the column was connected by tubing to a 4 liter suction flask used as a mixing vessel. The mixing vessel, operated with the aid of a magnetic stirrer, was filled initially with 4 liters of distilled water and was fed from a reservoir containing 40% (v/v) ethanol. Air pressure amounting to 3 psi was applied to the system at the reservoir. A pressure release in the form of a tube protruding into a column of mercury was placed before the reservoir inlet to prevent the total pressure from exceeding 10 psi. The system had been checked to assure it would withstand this pressure. The eluate was collected in 19 mL fractions using a fraction collector. After 250 tubes, the feed solution in the reservoir was changed to 10 liters of 50% (v/v) ethanol. A total of 400 tubes were collected. The flow rate decreased to 1 mL/min at the end of the separation.

Isolation of the Oligosaccharides

Colorimetric estimation of the carbohydrate content was used to locate the oligosaccharide components. Six mL aliquots of color reagent composed of 0.2% orcinol in 70% (v/v) sulfuric acid were added to 0.1 mL aliquots of every fifth

tube of eluate and the mixtures were heated in a steam bath for 20 minutes. The solutions were allowed to cool and absorbance and optical density at 550 nm were measured using a Bausch and Lomb Spectronic 20 Colorimeter. The color producing reaction is shown in Fig. 50. The heat and sulfuric acid break down the oligosaccharides to glucose which undergoes rearrangement to yield a furfural. The furfural then condensed with the orcinol to produce a colored product.

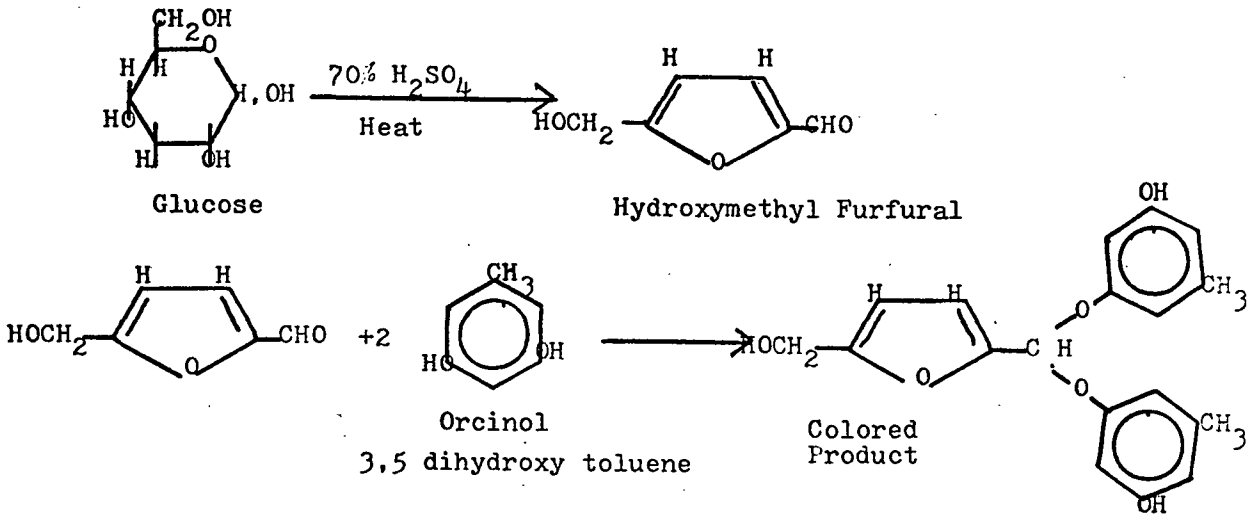


Figure 50. Color producing reaction used for locating the oligosaccharide components.

It was possible to identify the components of the test tubes by plotting optical density at 550 nm versus tube number. The components were identified and pooled as indicated in Fig. 51. The fractions were evaporated to dryness at reduced pressure on a rotary evaporator. A single separation yielded between 0.2 and 0.4 g of each cellodextrin, cellotriose through cellohexaose.

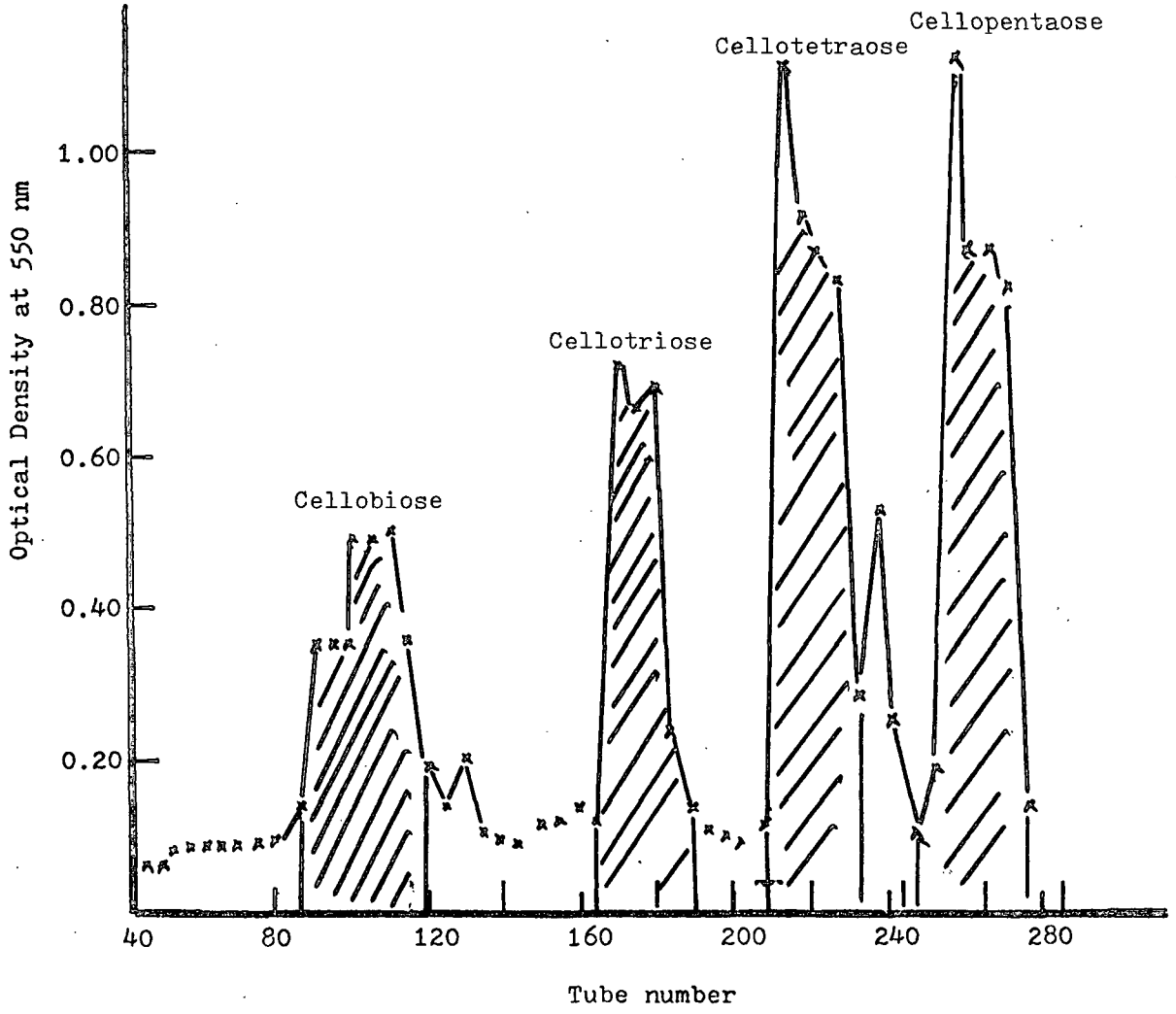


Figure 51. Optical density versus tube number - cellobiose through cellopentaose.

Recrystallization of the Cellodextrins

The cellodextrins were recrystallized by a method suggested by Dr. Chanzy.[58] The method consists of allowing acetone to diffuse into an aqueous solution of a particular cellodextrin. The cellodextrins are insoluble in acetone and precipitation occurs as the acetone concentration increases. A desiccator was used for the diffusion as illustrated in Fig. 52. This process was also found successful for recrystallizing xylobiose.[59]

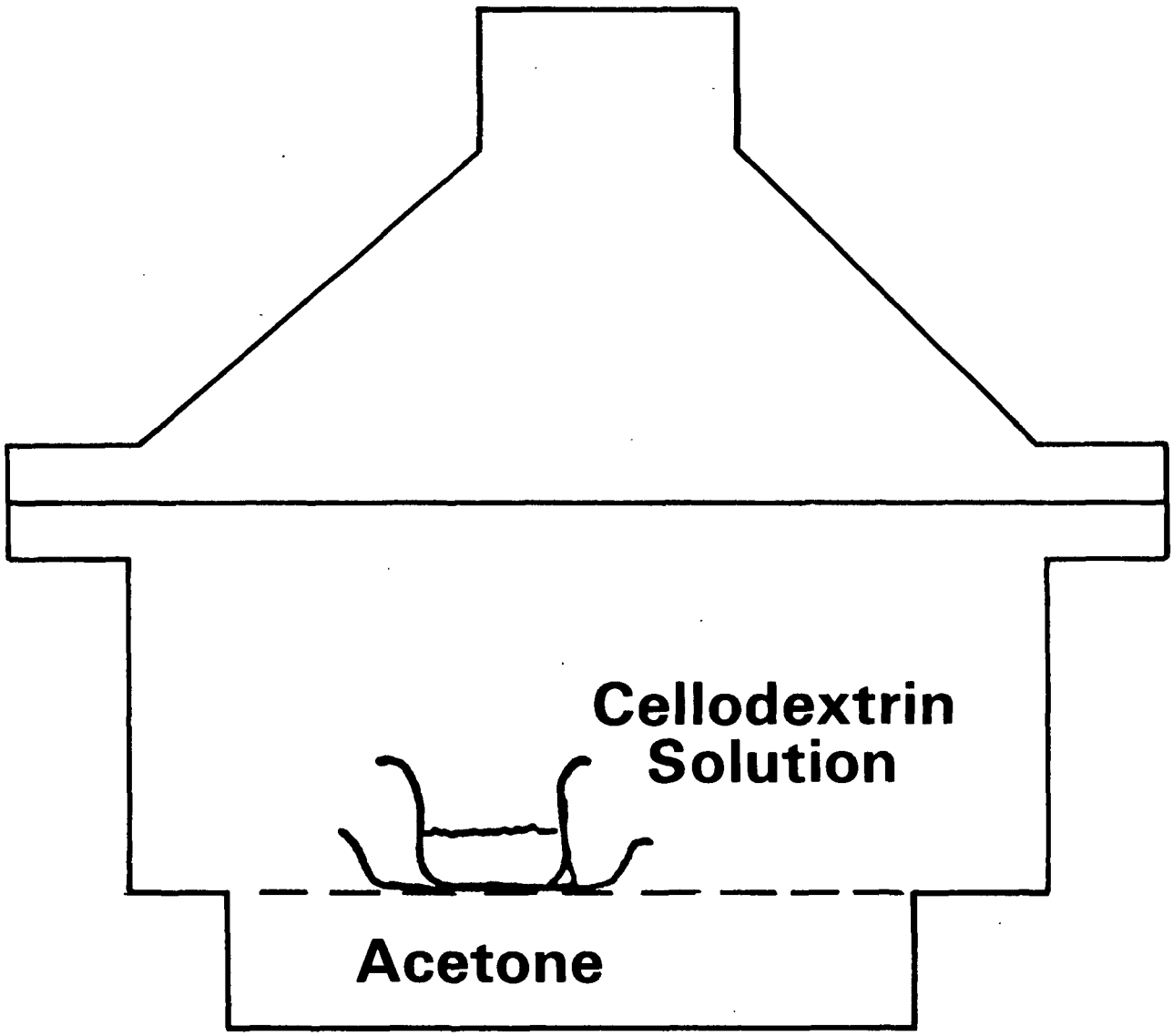


Figure 52. Apparatus for recrystallization of the cellodextrins.

Identification of the Cellodextrins

The purity of the cellodextrin samples was evaluated by a gas chromatographic method developed by J. Martin MacLeod.[60] The method consists of glc identification of the TMS derivatives of the cellodextrins on a 2 mm x 2 m OV 17 column. The TMS derivatives were produced using the Trisyl concentrate. Chromatograms were obtained for commercial samples of glucose and cellobiose and

the cellodextrin samples from cellobiose to cellotetraose. Cellopentaose and cellohexaose could not be detected by this method using the available equipment. The first sample of cellotriose contained an estimated 10% cellobiose. A second sample of cellotriose showed no cellobiose contamination and this second sample was used in acquisition of the observed spectra.

Observed Spectra

The observed spectra for the cellodextrins consisted of the infrared and the Raman spectra. The infrared spectra were recorded on a Perkin Elmer 621 Infrared Spectrophotometer. The Raman spectra were recorded on a Jobin Yvon Raman Spectrophotometer interfaced with a Tracor Northern TN-1500 minicomputer for data acquisition and analysis. The observed vibrational spectra for the cellodextrins are illustrated in Fig. 53 through 60. The vibrational spectra of cellulose I and cellulose II are shown in Fig. 61 through 64.

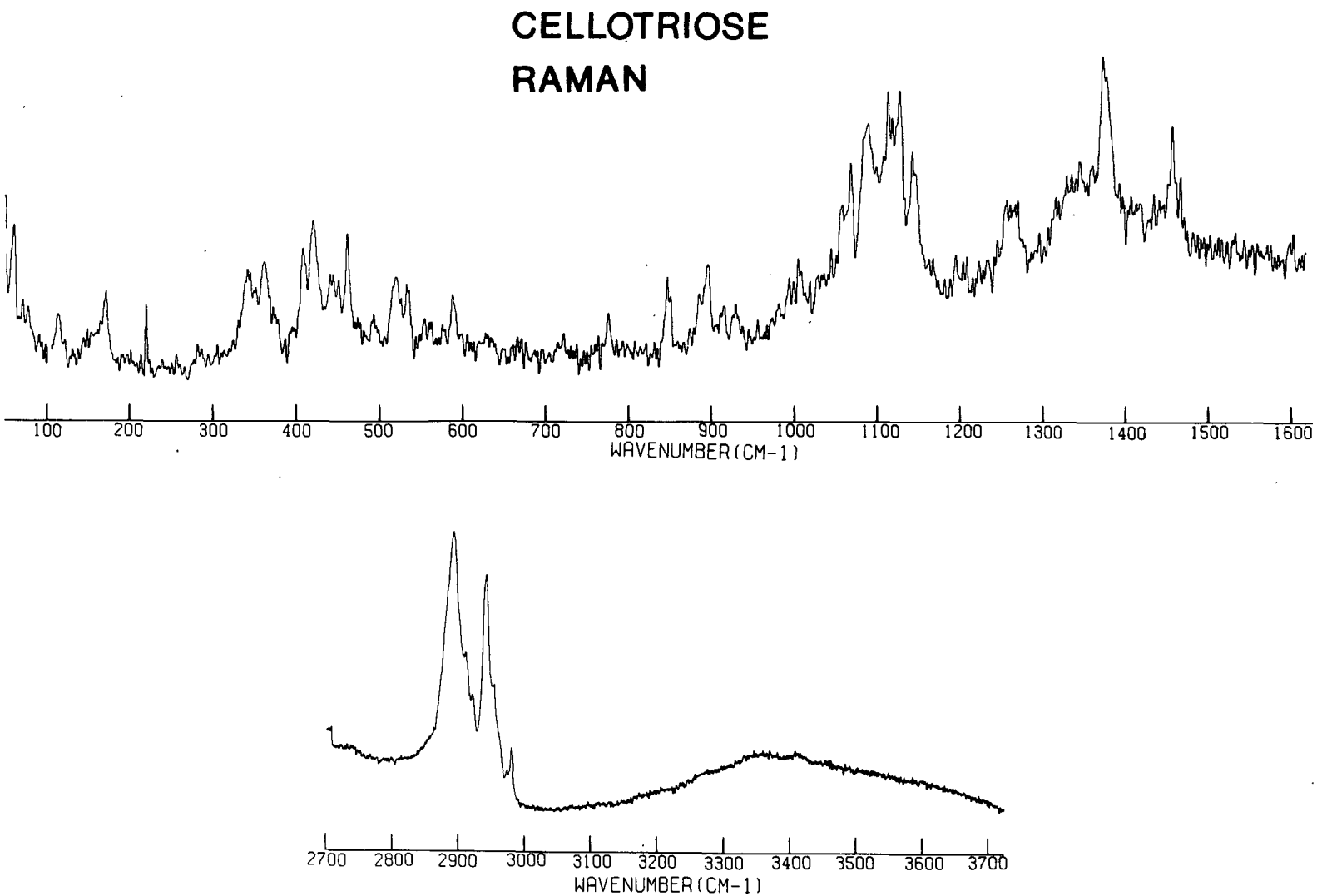


Figure 53. Raman spectrum of cellotriose.

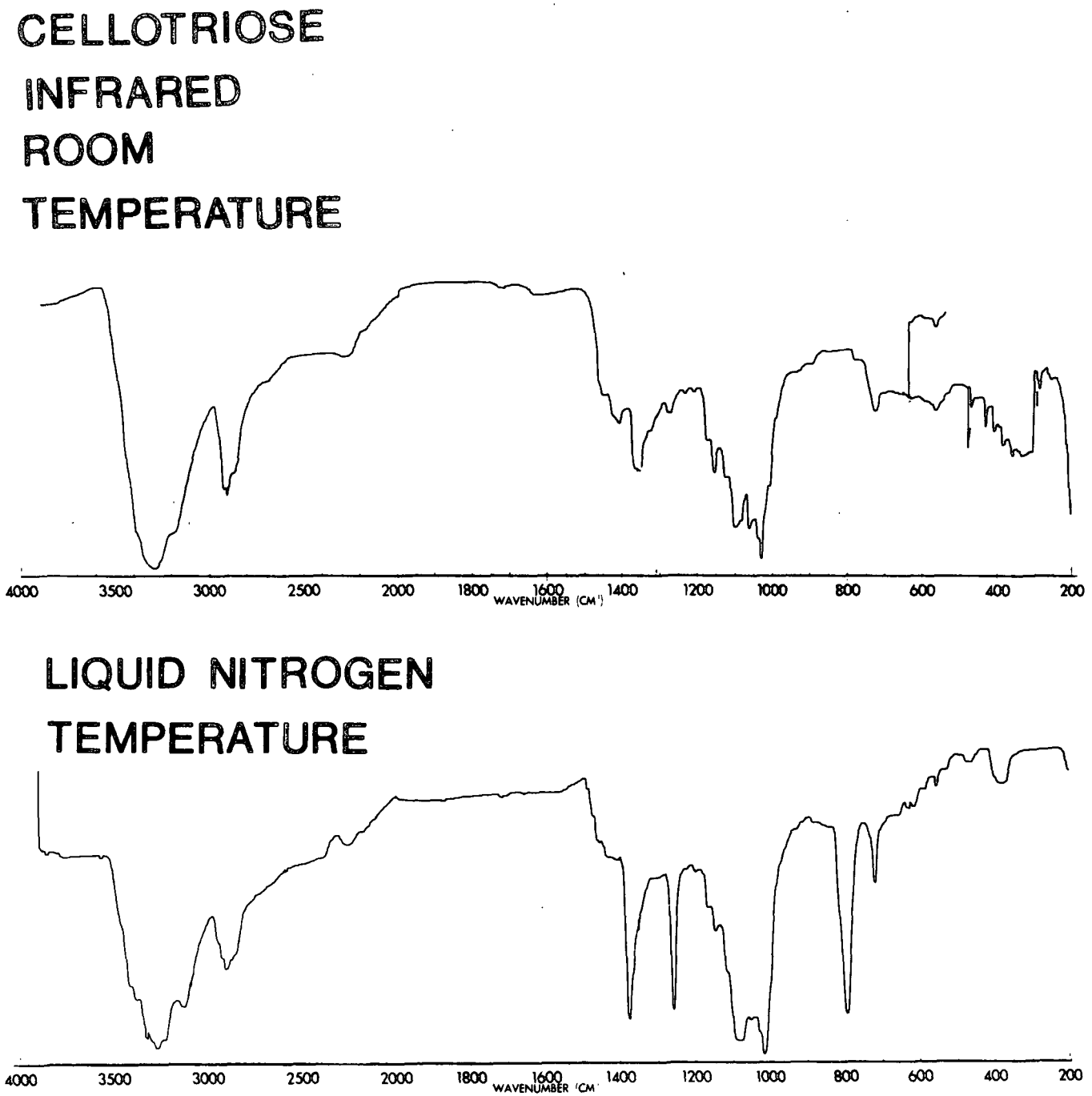


Figure 54. Infrared spectrum of cellotriose.

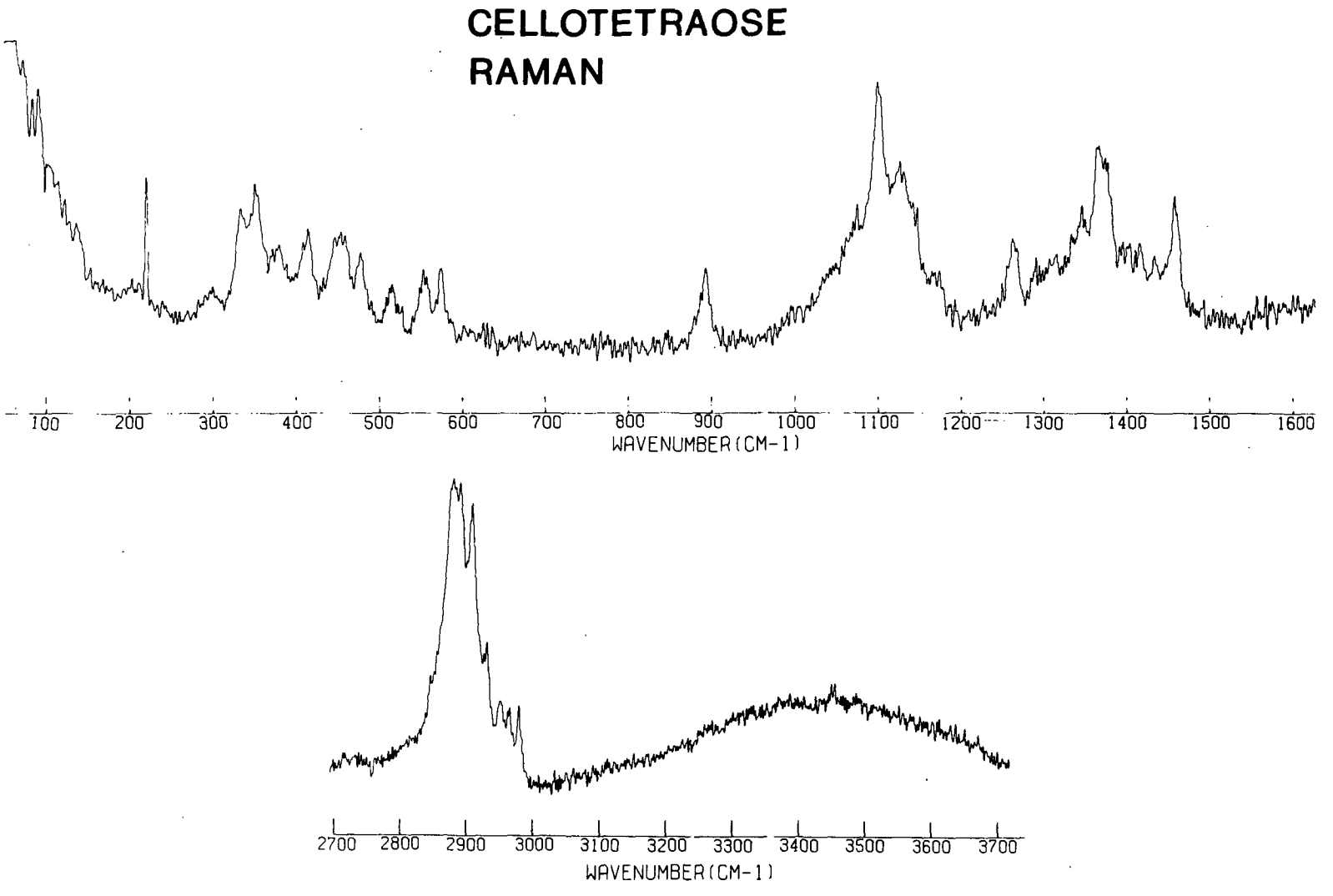


Figure 55. Raman spectrum of cellotetraose.

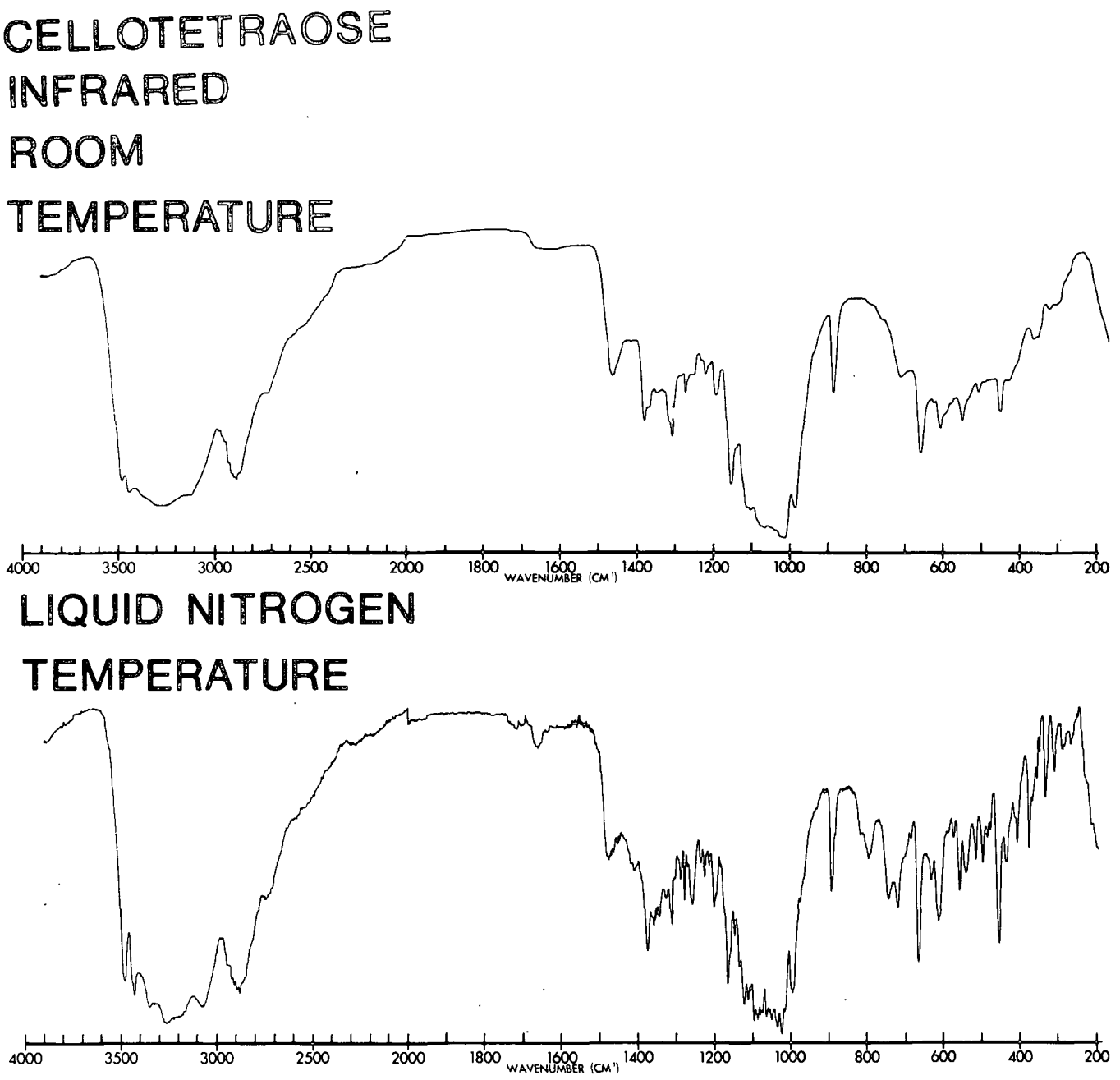


Figure 56. Infrared spectrum of cellotetraose.

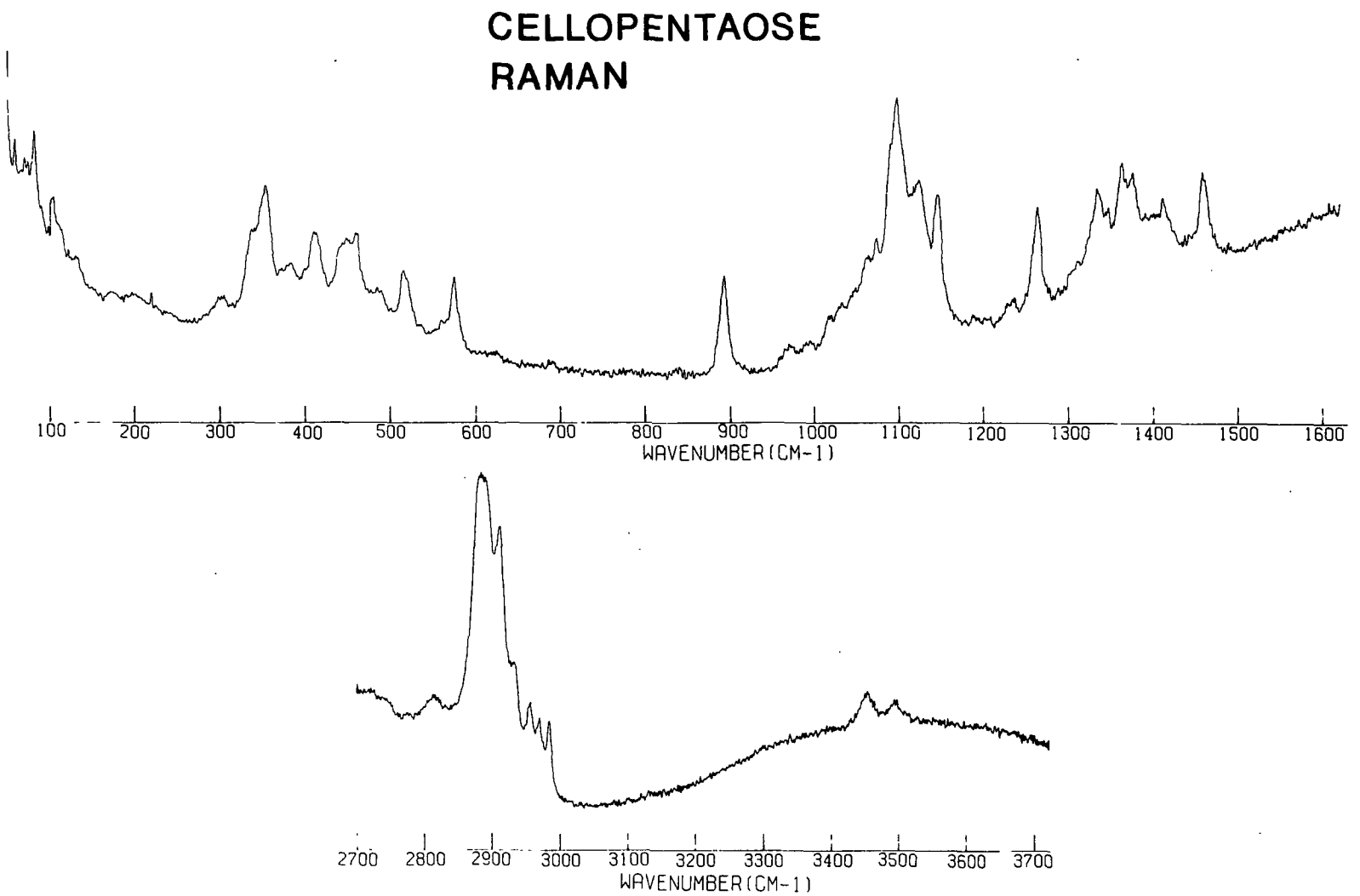


Figure 57. Raman spectrum of cellopentaose.

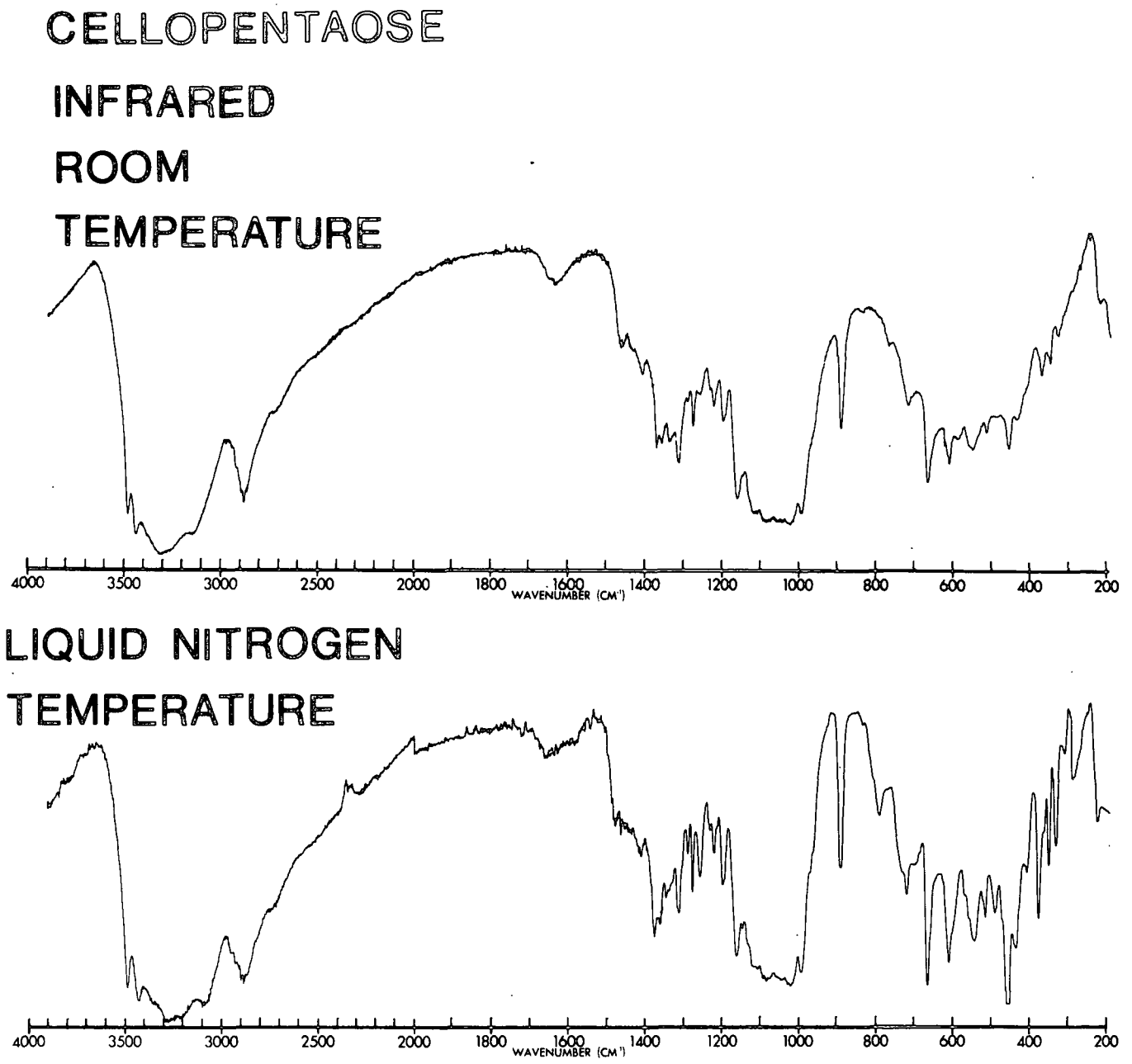


Figure 58. Infrared spectrum of cellopentaose.

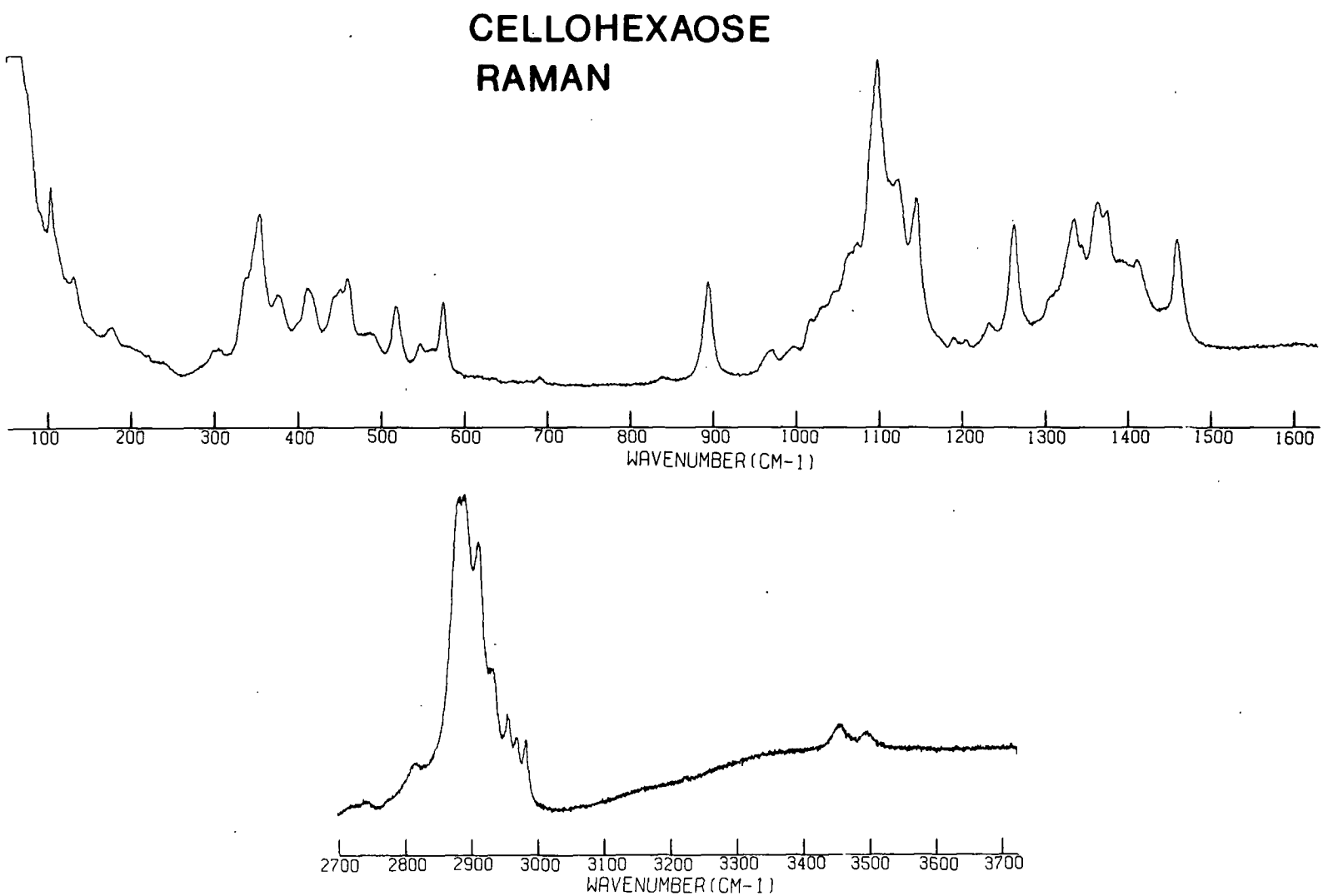


Figure 59. Raman spectrum of cellohexaose.

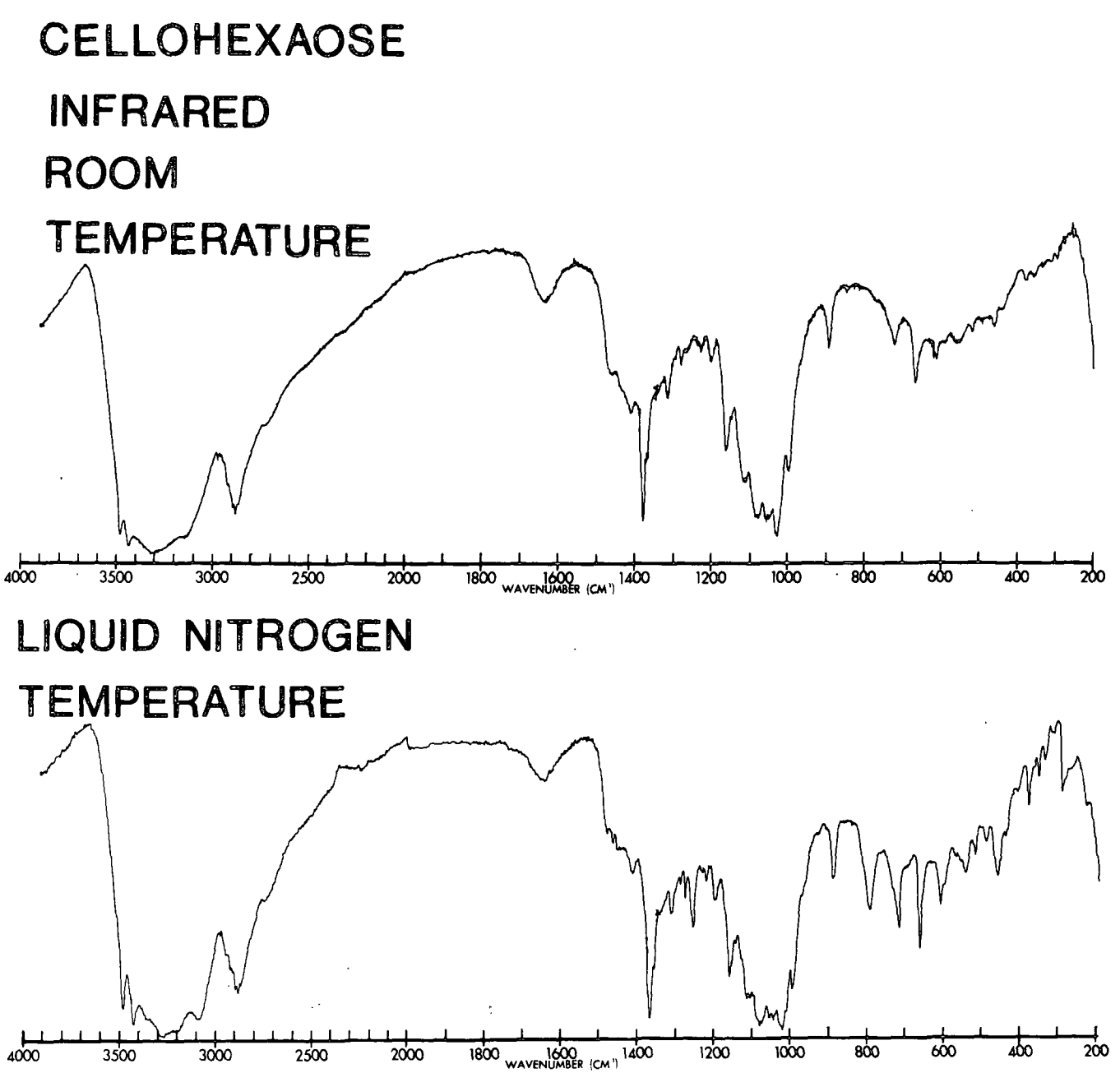


Figure 60. Infrared spectrum of cellohexaose.

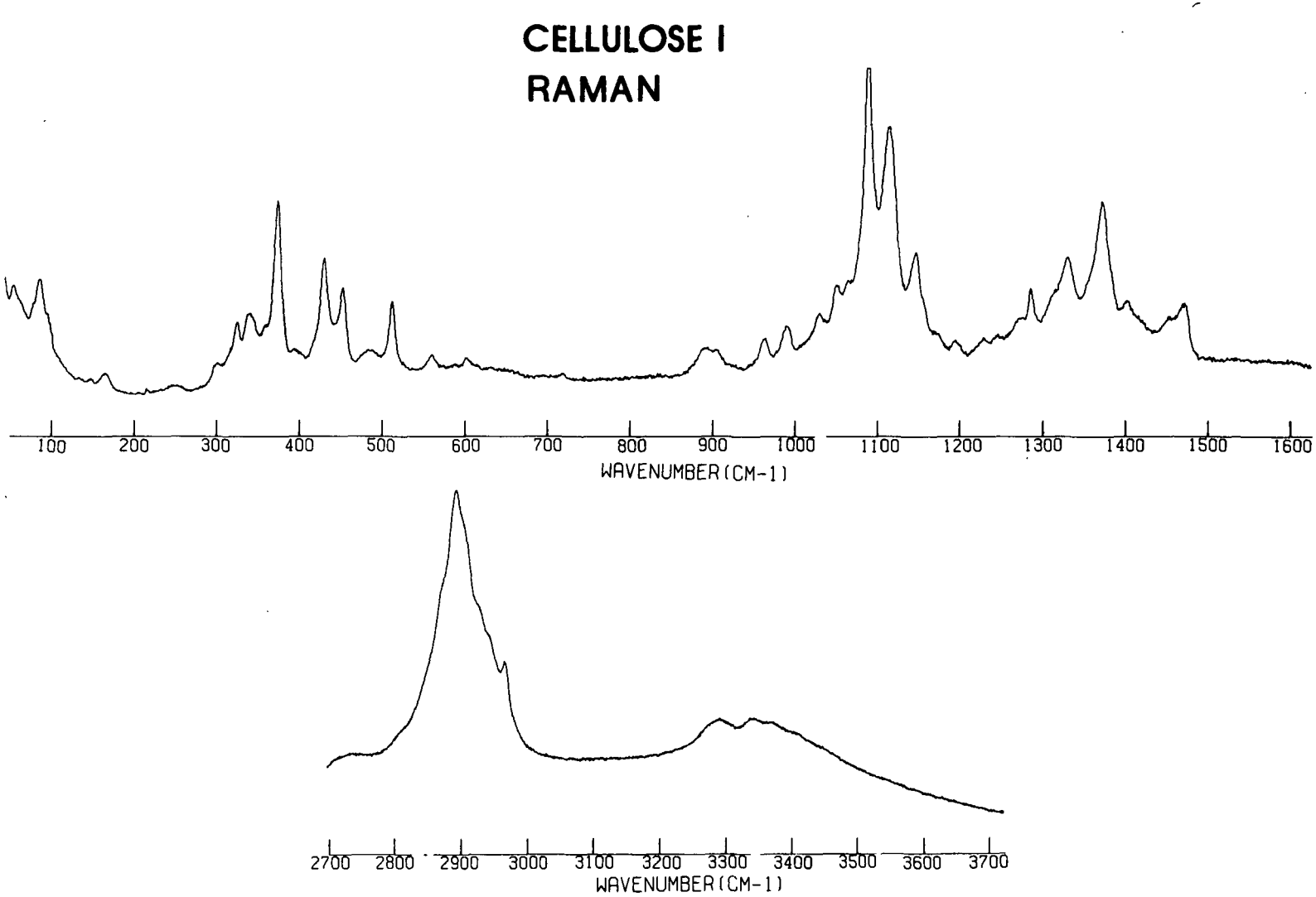


Figure 61. Raman spectrum of cellulose I.

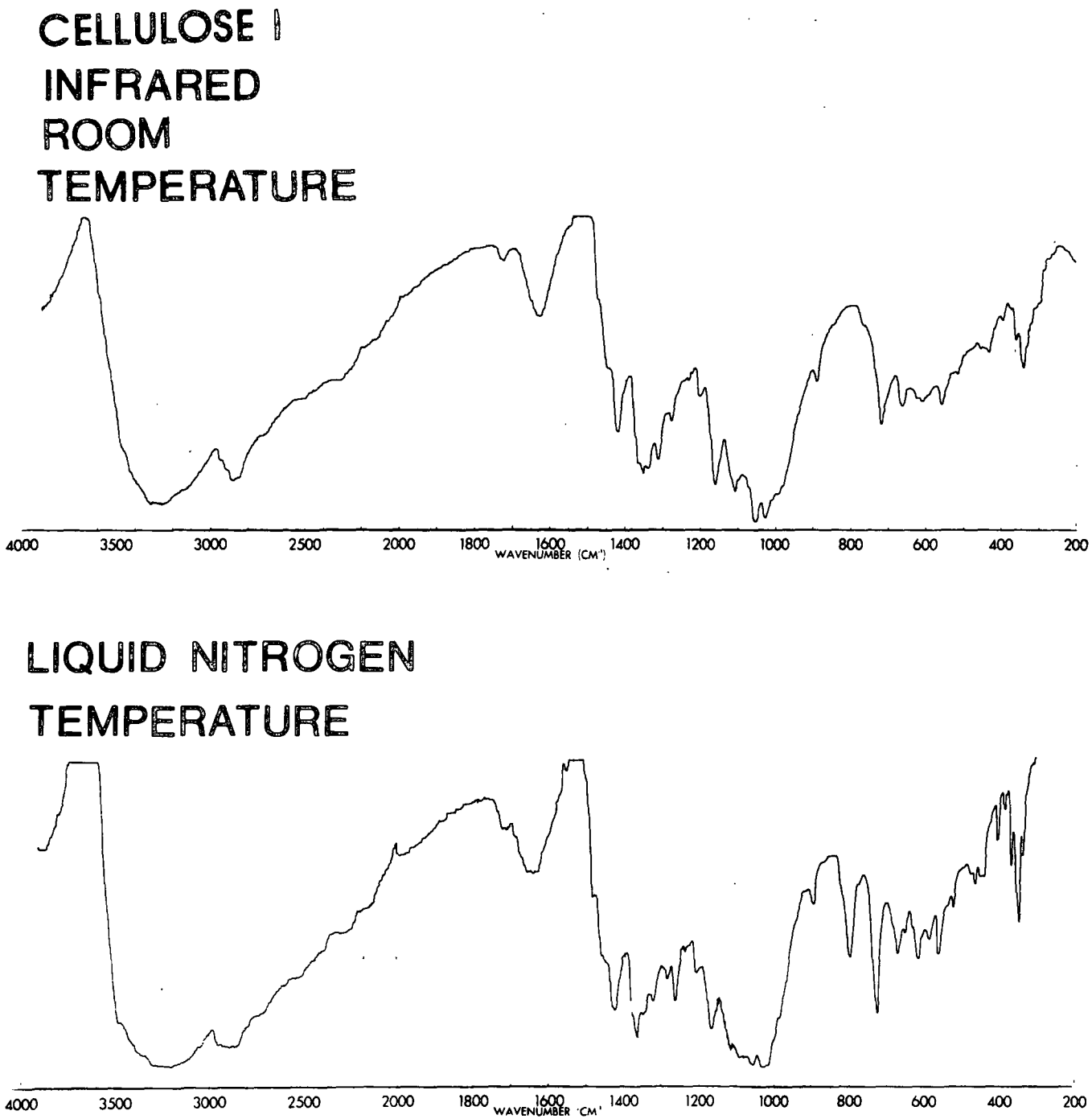


Figure 62. Infrared spectra of cellulose I.

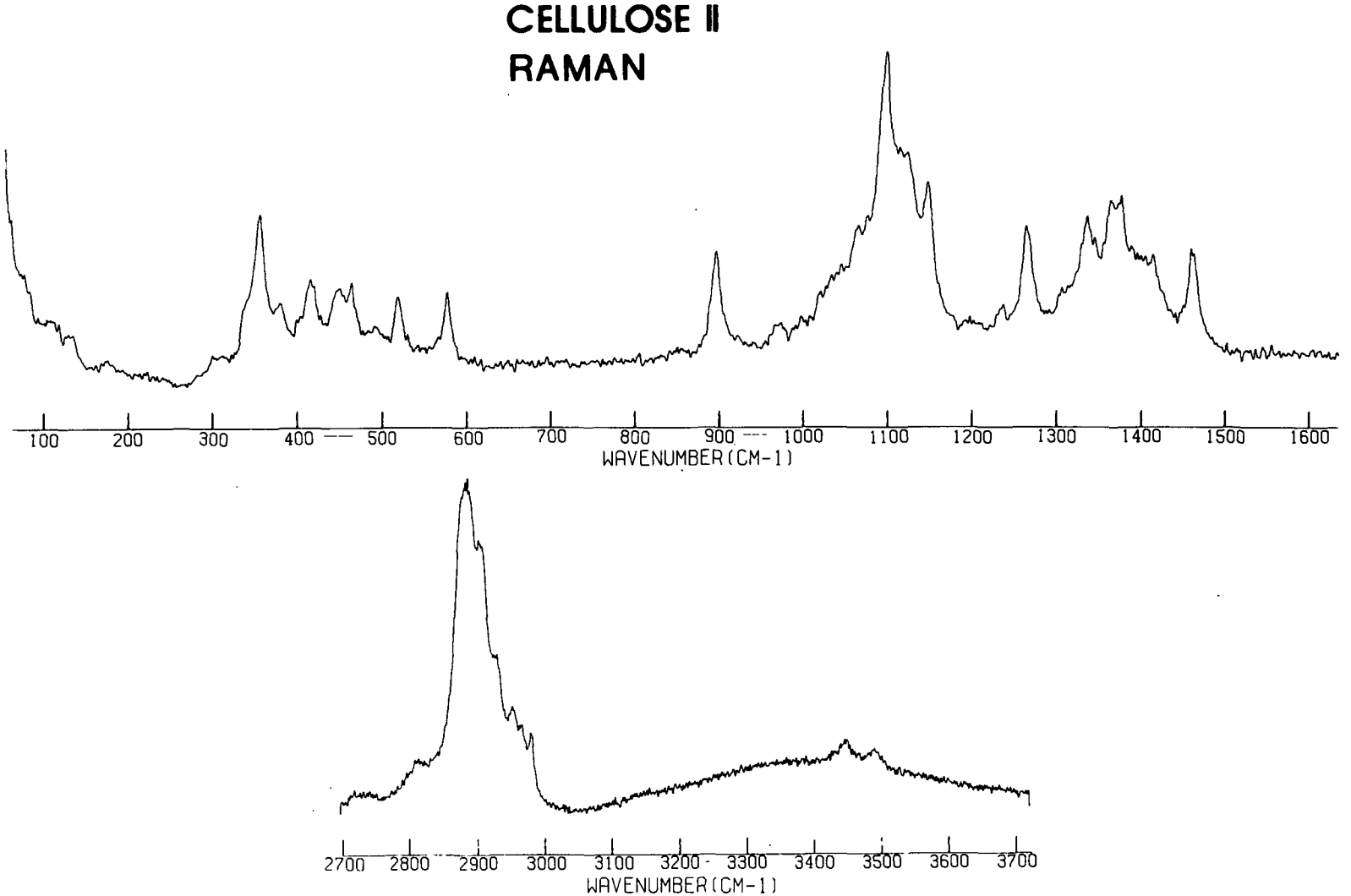


Figure 63. Raman spectrum of cellulose II.

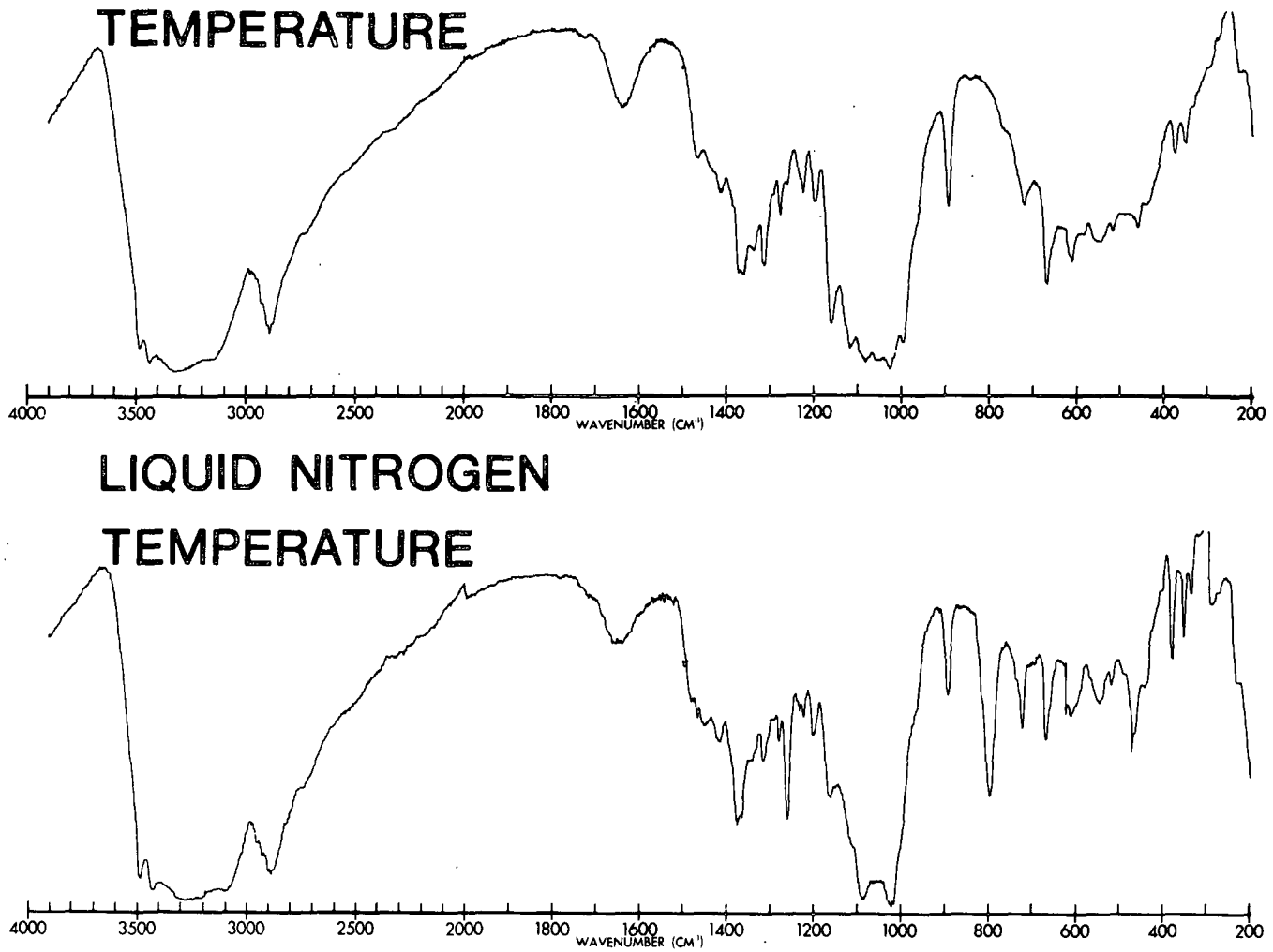


Figure 64. Infrared spectrum of cellulose II.

DESCRIPTION OF THE CELLOTETRAOSE STRUCTURES FOR THE CALCULATIONS

The Raman spectra for the cellodextrins and cellulose II below 1600 cm^{-1} are presented in Fig. 65 for direct comparison. It can be seen that there are only slight differences in the spectra for the oligomers cellotetraose and higher. Therefore, an interpretation of the vibrational spectra of cellotetraose based on normal coordinate analysis should also be applicable to the higher cellodextrins. Two cellotetraose structures were used for the normal coordinate analysis for two reasons. First, there is not a published crystal structure of cellotetraose to define a structural model to be used in the calculations. Second, the use of two structures allows the evaluation of the effect of conformational changes about the glycosidic linkage upon the calculated frequencies for cellotetraose. This can be a test of the conclusions obtained in the cellobiose calculations. The two structures were based on the atomic positions within the nonreducing ring in cellobiose with the dihedral angles about the glycosidic linkages set as in either cellobiose or methyl β-cellobioside. The first structure, with the dihedral angles at the glycosidic linkages set as in cellobiose, is illustrated in Fig. 66. Figure 67 illustrates the second cellotetraose structure with the dihedral angles about the bonds at the glycosidic linkage set as those found in methyl-cellobioside. The results of these calculations will be discussed separately in terms of a general interpretation of the vibrational spectra of the cellodextrins and the effect of conformational changes about the glycosidic linkages on the calculated frequencies for cellotetraose.

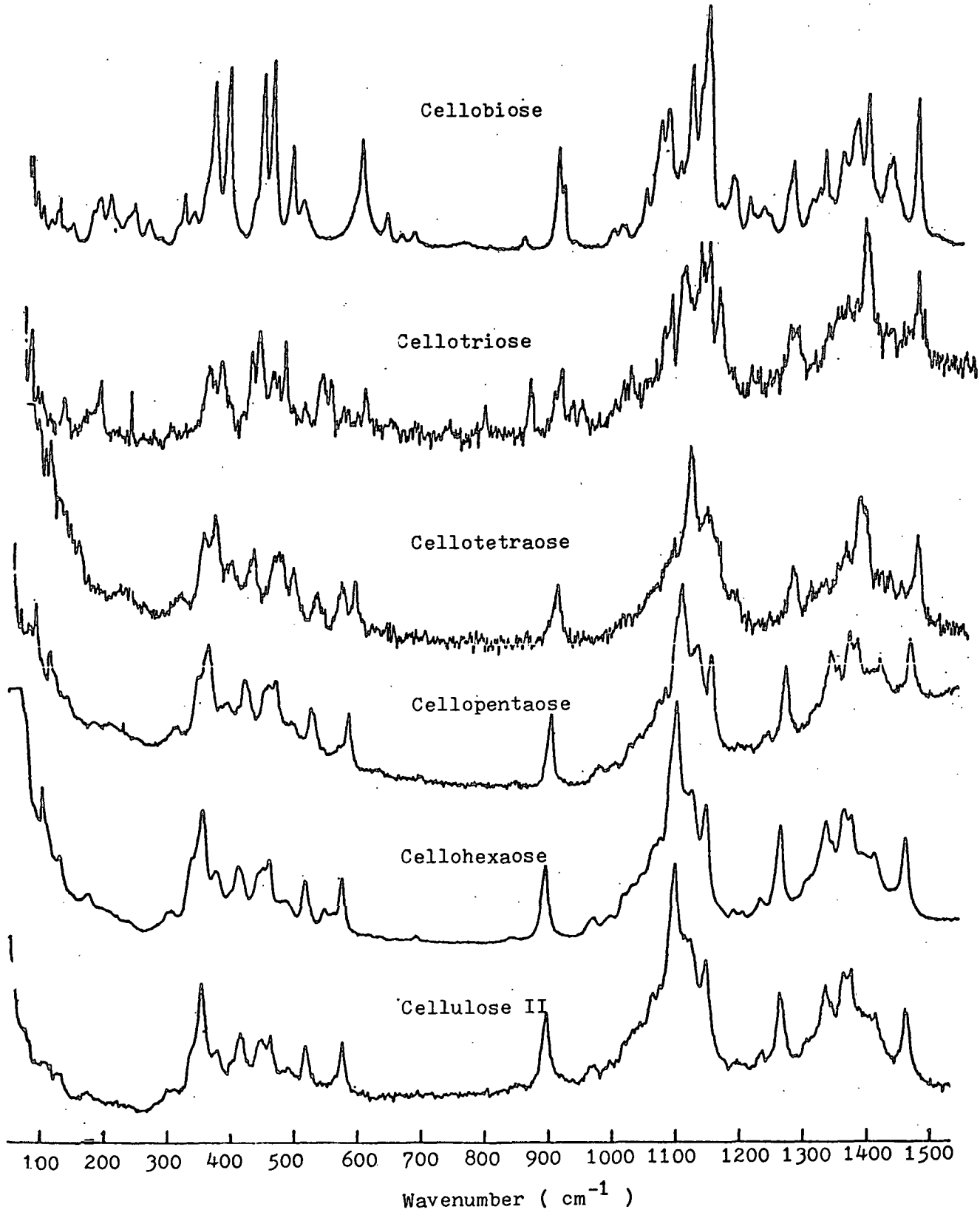


Figure 65. Comparison of the Raman spectra of the cellodextrins.

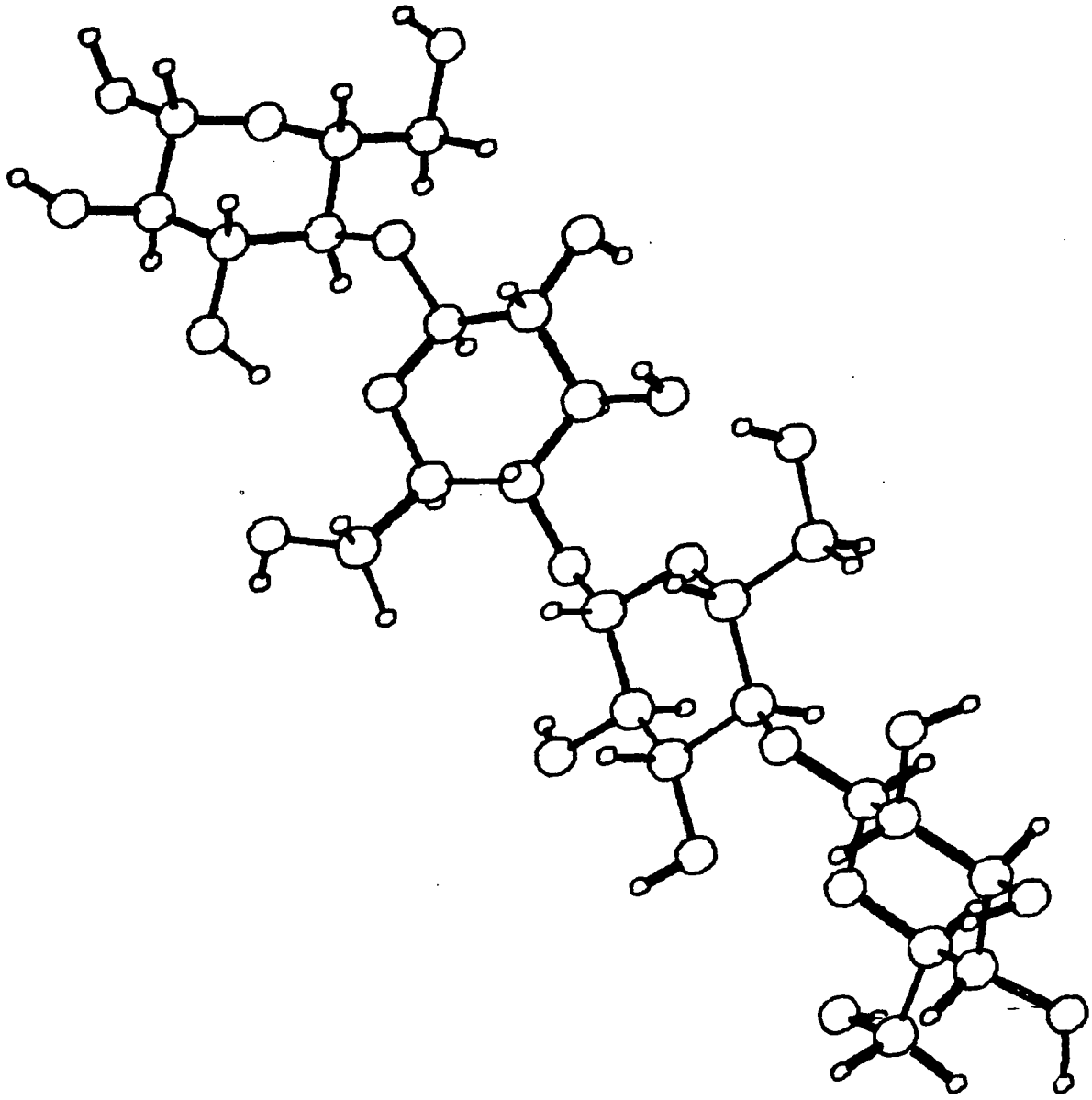


Figure 66. Illustration of the cellotetraose structure with linkages similar to cellobiose.

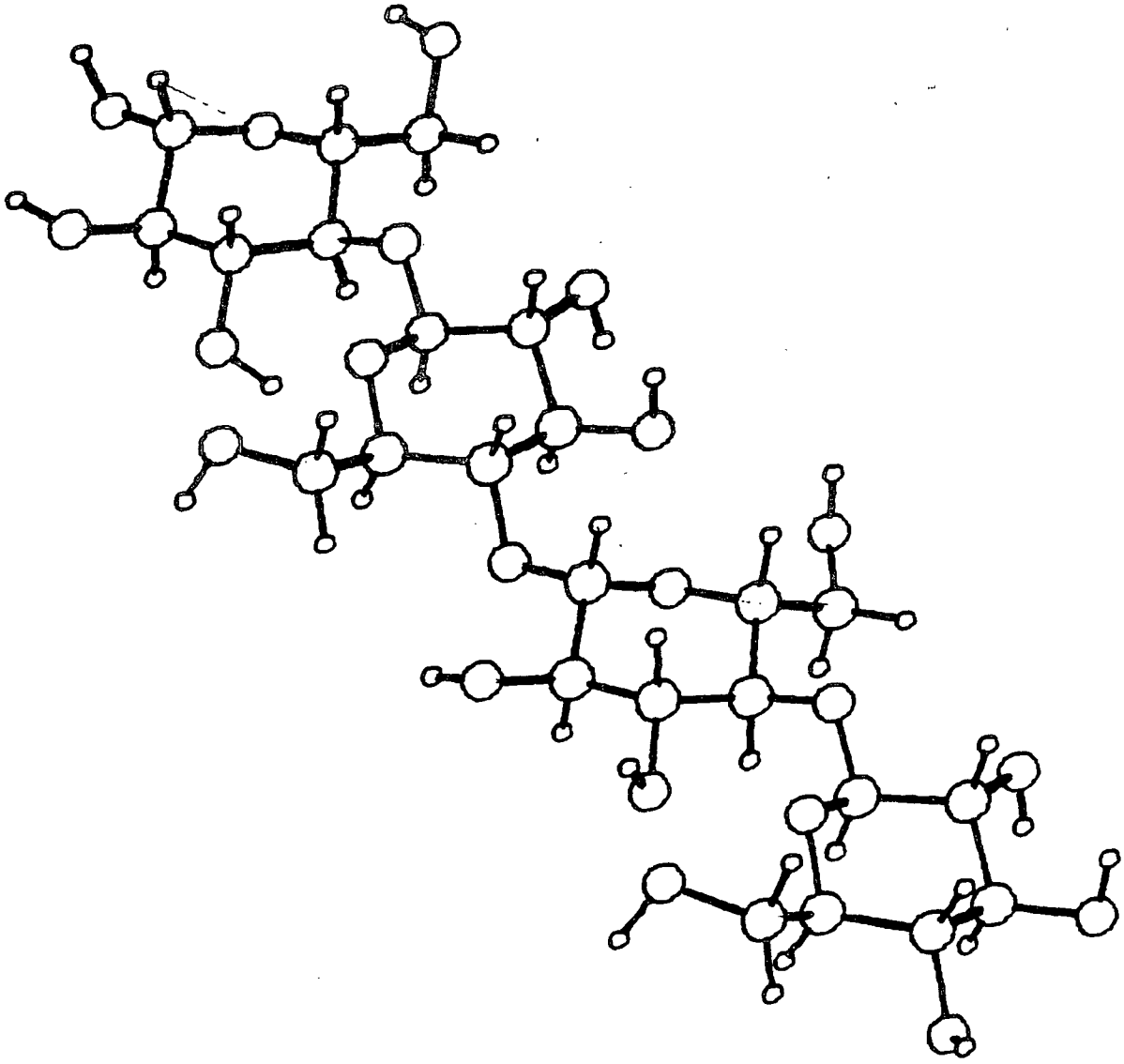


Figure 67. Illustration of cellotetraose structure with linkages as in methyl β-cellobioside.

INTERPRETATION OF THE VIBRATIONAL SPECTRA OF THE CELLODEXTRINS

The vibrational spectra of the cellodextrins may be divided into six sections for the purpose of discussion as was done for the disaccharides (see Section II). The basis for this interpretation, the normal coordinate calculations for cellotetraose, shows that the atomic motions are very similar to those calculated for the disaccharides for every region of the vibrational spectra. The first region is 3600-2650 cm^{-1}. This region is composed of the O-H and C-H stretching motions for cellotetraose. The interpretation of the O-H stretching region, 3600-3000 cm^{-1}, is presented in Section IV. The C-H stretching motions account for the strong observed bands between 2950 and 2800 cm^{-1} in the cellodextrins. The region 1500-1150 cm^{-1} is composed of methine (C-C-H and O-C-H), methylene (H-C-H) and hydroxyl (C-O-H) bending motions. The contributions to the third region, 1150-850 cm^{-1}, are primarily the C-O and C-C stretching motions with the bands calculated in the region 890-850 cm^{-1} dominated by the methine bending motions localized at the hydroxymethylene positions. The region between 850 and 400 cm^{-1} is constituted of heavy atom (O-C-C, O-C-O, and C-O-C) bending motion with some heavy atom stretching. The fifth region, 400-300 cm^{-1}, is dominated by the C-C-C and C-C-O bending motions. This region has calculated frequencies which are represented by significant motions of the atoms at the glycosidic linkages in cellotetraose as well as the disaccharides. Below 300 cm^{-1} the motions are predominantly torsional motions with some heavy atom bending. The detailed discussions of these regions for the disaccharides also apply to the individual regions of the calculated cellotetraose frequencies. The potential energy distributions greater than 2% are listed in Section IV for the cellotetraose calculations. Figure 68 compares the calculated frequencies for both cellotetraose structures with a representation of the observed Raman

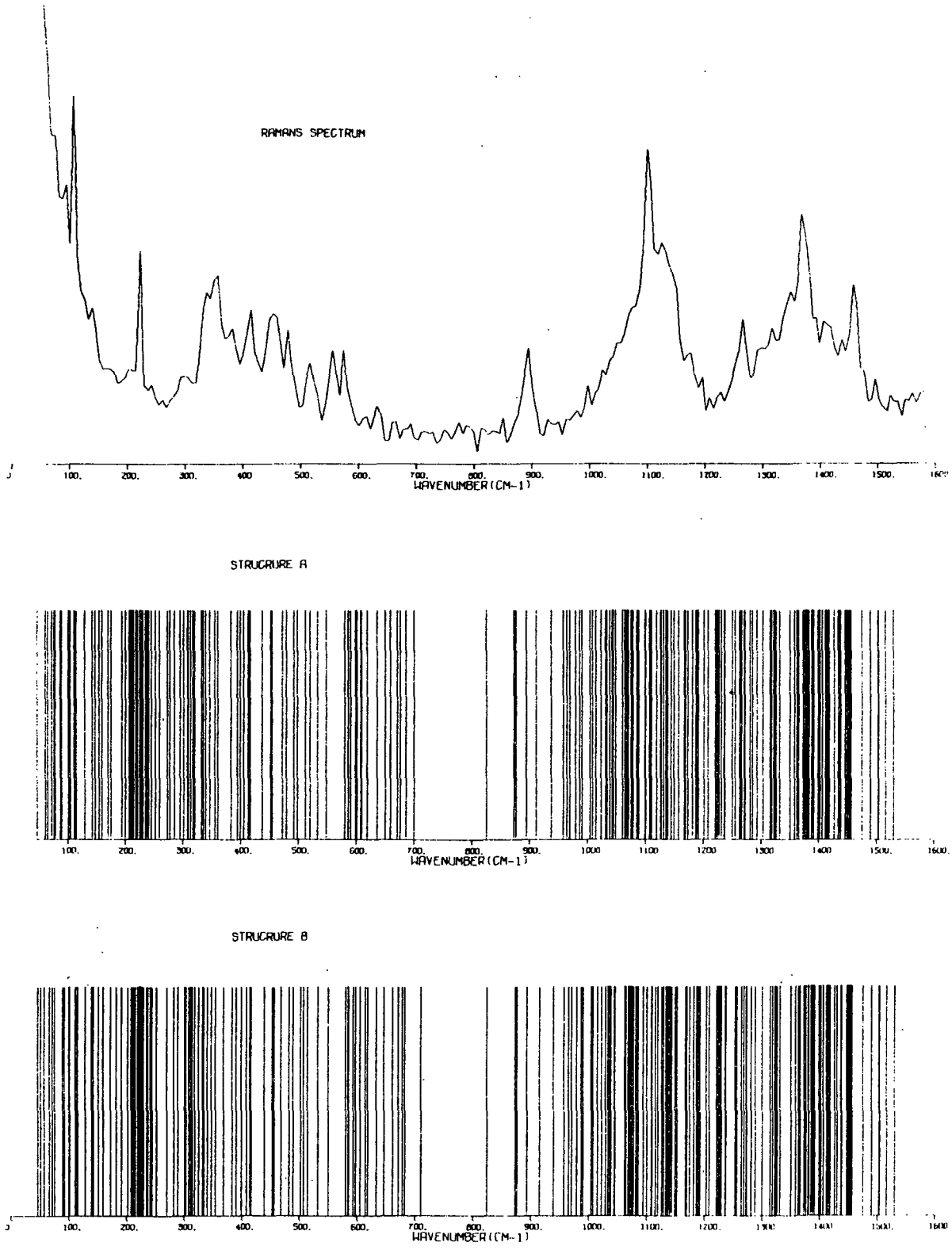


Figure 68. Comparison of calculated frequencies for two cellotetraose structures with a representation of the cellotetraose Raman spectrum.

spectrum of cellotetraose. The representation is a computer plot constructed from data values taken at 6.4 cm^{-1} intervals.

It can be seen in Fig. 68 that the general trends of the observed cellotetraose frequencies are reproduced by the calculated frequencies. The number of calculated frequencies prohibits individual assignments of the calculated frequencies to the observed bands.

EFFECT OF CONFORMATIONAL CHANGE AT THE GLYCOSIDIC LINKAGE ON THE CALCULATED FREQUENCIES OF CELLOTETRAOSE

The two structures used in the calculations for cellotetraose are illustrated in Fig. 66 and 67. Structure A, shown in Fig. 66, has all three linkages as found in cellobiose. Structure B, shown in Fig. 67, has all three linkages as found in methyl β-cellobioside. The calculations on various cellobiose conformations indicated that a structure similar to structure B should be found in cellulose I and a structure similar to structure A should be found in cellulose II. The calculations for cellotetraose were done, in part, to test these conclusions from the cellobiose calculations.

The calculated frequencies for these two structures are shown in Fig. 68 in comparison with a representation of the Raman spectrum of cellotetraose. There were many slight differences between the calculated frequencies for the two structures. There was only one significant difference between the calculated frequencies for structures A and B. This was in the frequency range 400-300 cm^{-1}. This difference reflects the observed difference between the observed spectra of cellulose I and cellulose II. Furthermore, it is in agreement with the prediction made from the cellobiose calculations that structure A would be similar to cellulose II, and structure B would be similar to cellulose I. This

has been taken as support for the conclusions made concerning the cellobiose calculations. Those conclusions were that the observed vibrational spectra of cellulose I and cellulose II can be accounted for in terms of conformational differences at the glycosidic linkages. Based on these findings, both the cellobiose and cellotetraose calculations, the cellulose I structure is believed to have linkages similar to that found in methyl β-cellobioside and cellulose II is believed to have linkages similar in conformation to that found in cellobiose.

SECTION IV

INTERPRETATION OF THE O-H STRETCHING REGION FOR SEVERAL DISACCHARIDES, THE CELLODEXTRINS, CELLULOSE I, AND CELLULOSE II

INTERPRETATION OF THE O-H STRETCHING REGION OF THE VIBRATIONAL SPECTRA OF SEVERAL RELATED DISACCHARIDES

The O-H stretching region of the vibrational spectra, 3000-3600 cm^{-1} was investigated for cellobiose, methyl β-cellobioside, and α-lactose. These data, in conjunction with the crystal structures for these compounds, allowed the interpretation of this region of the vibrational spectra in terms of intermolecular and intramolecular hydrogen bonding. This interpretation has been applied to xylobiose, β-lactose, and sophorose, none of which have a known crystal structure. The interpretation has also been applied to maltose, the 1-4 linked disaccharide of glucose.

The vibrational spectra of the carbohydrates are very complex when taken as a whole. However, certain regions of the vibrational spectra may be adequately interpreted by the application of the group frequency approach. One such region is the O-H stretching region, particularly for the disaccharides. It should be noted that the group frequency approach is only qualitative and does not constitute a definitive description of the motions of this region.

Free O-H stretching vibrations would be expected to have associated frequencies between 3650 and 3800 cm^{-1} such as the 3656 and 3755 cm^{-1} bands for water vapor. Participation in a hydrogen bond alters the nature of the O-H bond and thus changes the frequency associated with the O-H stretching motion. It has been well documented[61] that the O-H stretching vibration is shifted to a lower frequency according to the strength of the hydrogen bond. Very strong hydrogen bonds may shift the characteristic O-H stretching vibration to a value as low as

600 cm^{-1}. The hydrogen bonds associated with the carbohydrates, although very important, rarely shift the O-H stretching frequency to values lower than 3000 cm^{-1}.

The shape of the band associated with a particular hydrogen bonded O-H stretching vibration may be interpreted in terms of the coupling of the O-H stretching motion with other motions through the hydrogen bond. The interpretation presented here will be entirely qualitative and no attempt will be made to evaluate such things as the anharmonicity of the O-H stretching which would require careful observation of overtone bands which was not possible in this study.

The vibrational spectra of cellobiose, methyl β-cellobioside, and α-lactose between 3000 and 3600 cm^{-1} may be interpreted in terms of their crystal structures.[18,32,33] Raman and infrared spectra for these compounds are included in Fig. 69. Crystal structures, denoting intramolecular hydrogen bonds are shown in Fig. 70. It can be seen that all three spectra show a broad band with very little definition between 3000 and 3200 cm^{-1}. This broad band is associated with the O-H stretching motions of hydroxyl units participating in intramolecular hydrogen bonds. These are the hydrogen bonds which account for many of the crystal effects and the cohesive forces in the solid samples. The region between 3300 and 3600 cm^{-1} shows sharp, well-defined bands in cellobiose and α-lactose. These bands are associated with the isolated intramolecular hydrogen bonds shown in Fig. 70 or with the hydrogen bond with the water of hydration in α-lactose. The observed vibrational frequencies and the O...O distances associated with the appropriate hydrogen bond are listed in Table 6 and compared to literature observations[62] in Fig. 71. It can be seen that these observations are consistent with the observations for hydrogen bonding in other studies. The

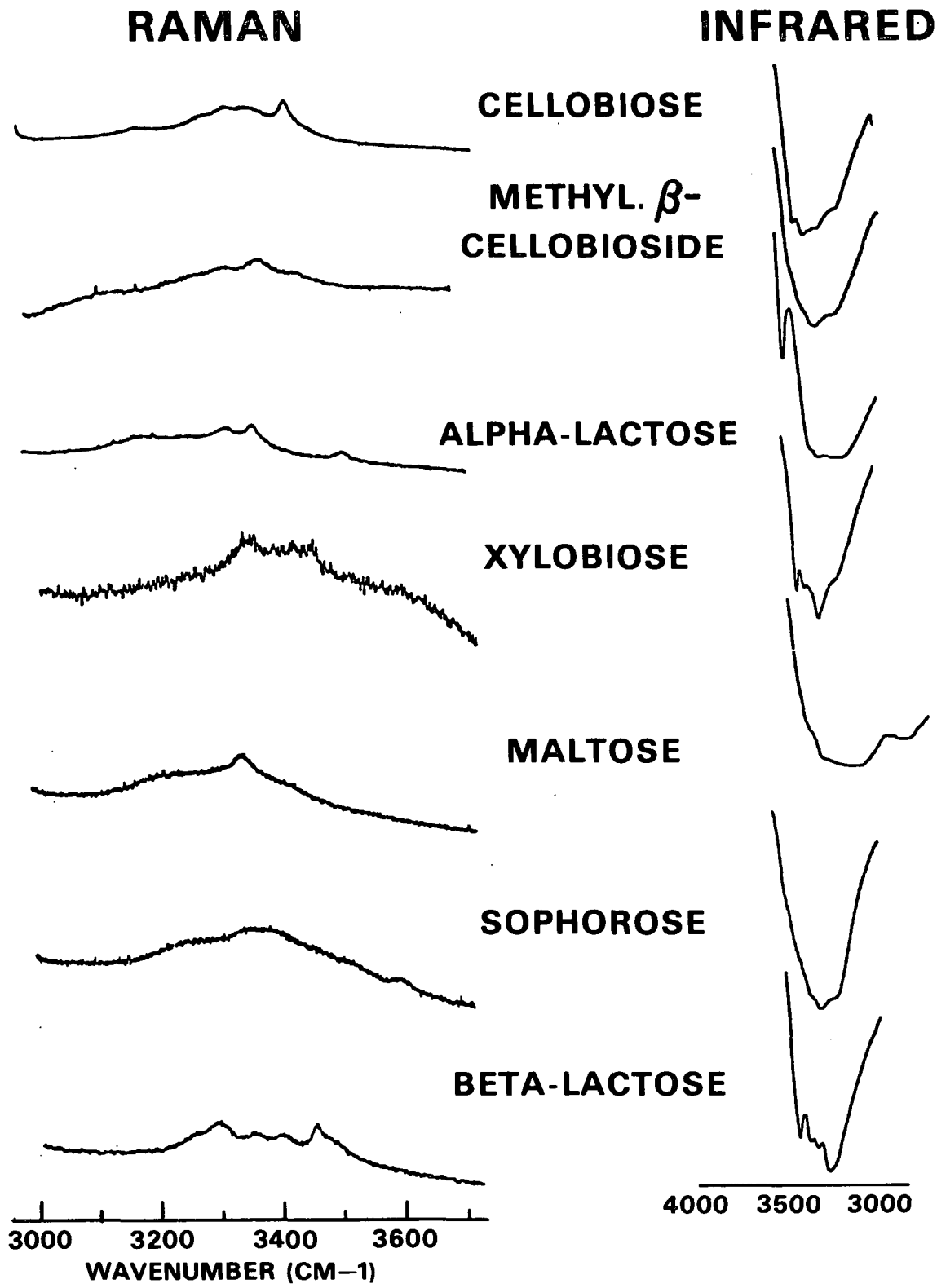
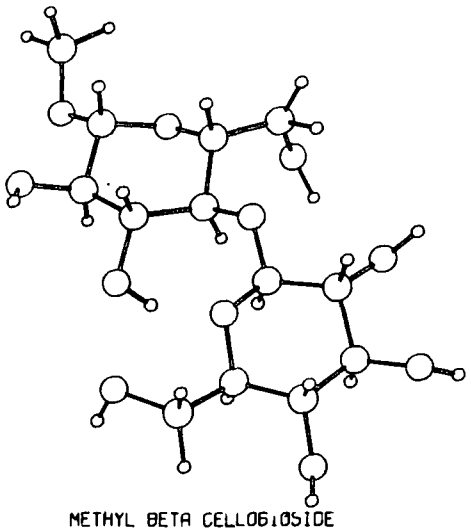
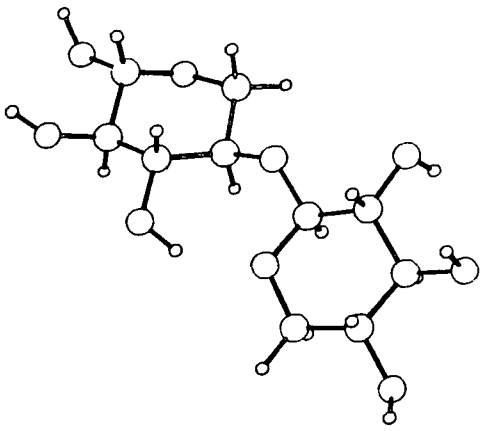


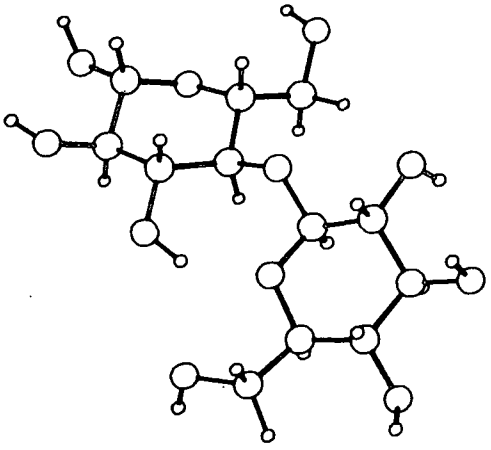
Figure 69. Raman and infrared spectra 3000-3600 cm^{-1} for cellobiose, methyl β-cellobioside, α-lactose, xylobiose, maltose, sophorose, and β-lactose.



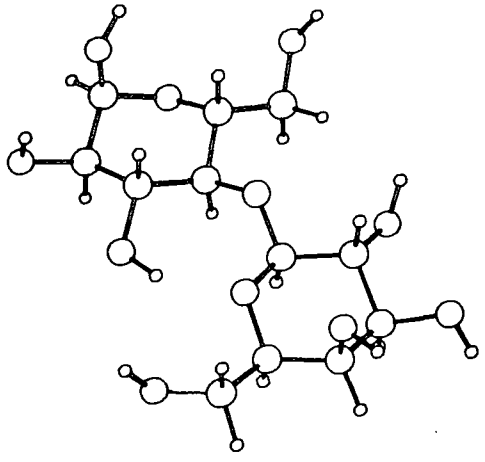
METHYL BETA CELLOBIOSIDE



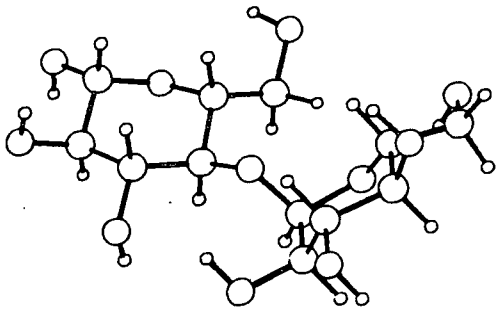
XYLOBIOSE STRUCTURE FOR CALCULATIONS



CELLOBIOSE CRYSTAL STRUCTURE



α - LACTOSE CRYSTAL STRUCTURE



MALTOSE STRUCTURE FOR CALCULATIONS

Figure 70. Structures of cellobiose, methyl β-cellobioside, xylobiose, α-lactose, and maltose illustrating possible hydrogen bonds.

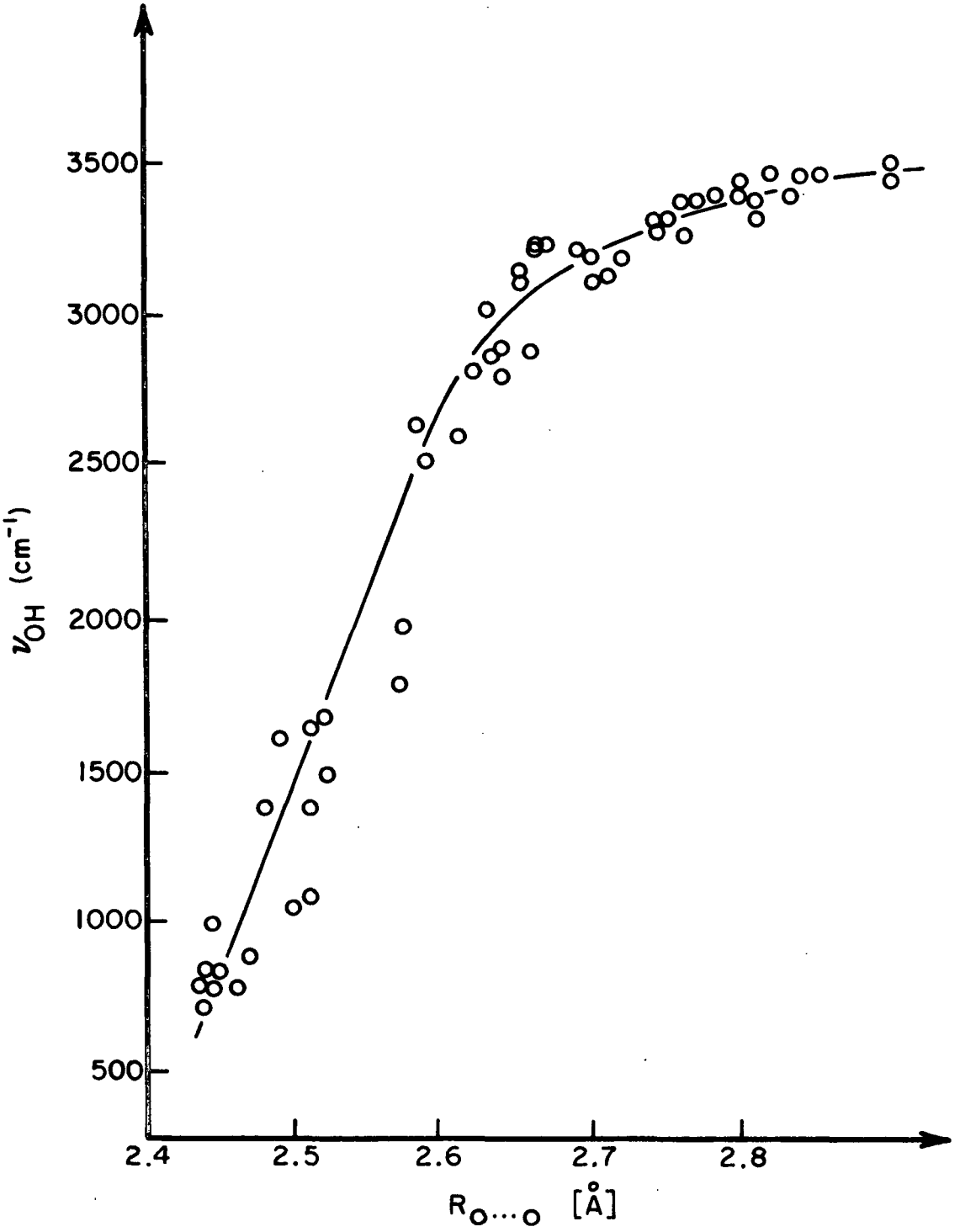


Figure 71. Comparison of frequency value with O...O distance.

broad band observed for methyl β-cellobioside in this region, 3300-3600 cm^{-1}, can be attributed to a combination of the effects of the bifurcated intramolecular hydrogen bond, illustrated in Fig. 70, and the intermolecular hydrogen bonds. The broadening of the band associated with that bifurcated bond may be due to either a very broad potential energy caused by the bifurcation or to the coupling of that hydroxyl motion with the motions of the hydroxylmethylene position through part of the bifurcated hydrogen bond. In either case, the bifurcation of the intramolecular hydrogen bond results in the broadening of the O-H stretching band associated with that hydrogen bond. Thus the O-H stretching region of the vibrational spectra may be interpreted in terms of intermolecular and intramolecular hydrogen bonds.

Table 6. Observed vibrational bands and the O...O distances associated with the hydrogen bonds in which they participate.

Molecule	Band (cm^{-1})	O...O (Å)
Cellobiose	3421	2.761
α-Lactose	3382	2.748
	3523	2.811
Methyl β-cellobioside	--	2.914
		2.762
Xylobiose	3357	--
β-Lactose	3290	--
	3448	--
Sophorose	--	--
Maltose (Methyl β-maltopyranoside)	3345	2.825[34]

This interpretation has been successfully applied to the O-H stretching regions of the spectra for xylobiose, β-lactose, and sophorose for which no crystal structures have been published. The observed spectra for these compounds in the region 3000-3600 cm^{-1} are illustrated in Fig. 69. The broad bands in the region 3000-3300 cm^{-1} are associated with the O-H stretching of the

hydroxyl units participating in the intermolecular hydrogen bonds present. The sharp band at 3357 cm^{-1} for xylobiose may be associated with the motions of the hydroxyl unit participating in the intramolecular hydrogen bond. The sharp bands for β-lactose may be associated with either an intramolecular hydrogen bond (3448 cm^{-1}) or a hydrogen bond including the water of hydration (3290 cm^{-1}). This is also observed in α-lactose. Figure 70 indicates the intramolecular hydrogen bonds for xylobiose and β-lactose. The frequencies associated with them are listed in Table 6. Any proposed structure for these compounds should be consistent with these observations. The vibrational spectrum of sophorose, the β 1-2 linked disaccharide of glucose, is shown in Fig. 69 also. It has a broad band in the O-H stretching region similar to that observed for methyl β-cellobioside. The implications of this observation are not clear at this time.

The Raman and infrared spectra for the O-H stretching region for maltose are also presented in Fig. 69. These spectra can also be interpreted in terms of intermolecular and intramolecular hydrogen bonds. Similar broad bands, associated with the hydroxyls participating in the intermolecular hydrogen bonds, are present in the region 3000-3300 cm^{-1}. No sharp bands appear in the infrared spectrum, but one sharp band appears in the Raman spectrum. This band may be associated with the intramolecular hydrogen bond shown for maltose in Fig. 70 or, more likely, with a hydrogen bond associated with a water of hydration. The crystal structure of methyl β-maltopyranoside[34] is presented in Fig. 72 showing the hydrogen bonds with the water of hydration for this molecule. This structure may be taken as the basis for the interpretation of the O-H stretching region of the maltose spectra. The sharp band at 3346 cm^{-1} in the maltose Raman spectrum may be associated with the stretching of the hydroxyl participating in an intramolecular hydrogen bond or with a hydroxyl participating in a hydrogen bond with

the water of hydration if maltose crystallizes as a hydrate. The latter is the more likely considering the frequency and the fact that maltose is supplied commercially as the hydrate.

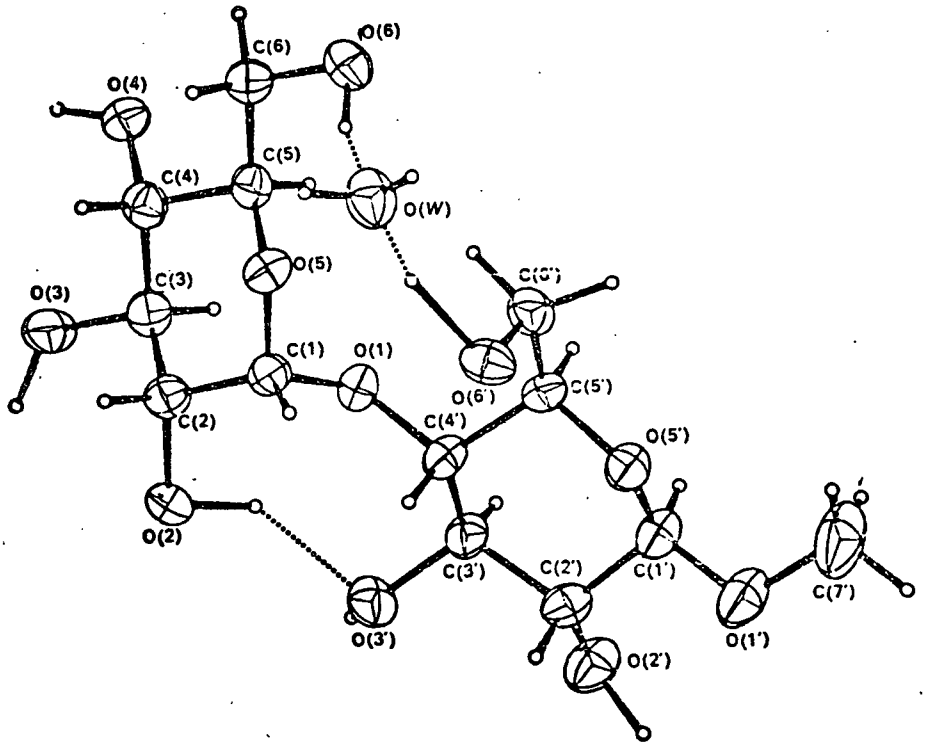


Figure 72. Crystal structure of methyl β-maltopyranoside.[34]

It has been shown that the O-H stretching region of several disaccharides may be interpreted in terms of the effects on the O-H stretching frequencies from intermolecular and intramolecular hydrogen bonds. Certain implications can be drawn from this region concerning the structure of compounds for which crystal structures have not been published based on the hydrogen bonding indicated by the O-H stretching region.

INTERPRETATION OF THE O-H STRETCHING REGION OF THE VIBRATIONAL SPECTRA OF THE CELLODEXTRINS

A very interesting effect is observed in the O-H stretching region of the vibrational spectra of the cellodextrins as one progresses to the larger molecules. The region 3000-3600 cm^{-1} of the vibrational spectra of the cellodextrins is presented in Fig. 73. The preceding section demonstrated that the O-H stretching region of the vibrational spectra of cellobiose indicates the presence of a single isolated intramolecular hydrogen bond. The Raman and infrared spectra of cellotriose indicate that there is no isolated intramolecular hydrogen bond. Cellotetraose shows the presence of two distinct bands in the infrared and indications are also present in the Raman of two distinct bands. The higher cellodextrins and cellulose II show two distinct bands in both the infrared and Raman. The presence of these two bands indicate that there are two isolated intramolecular hydrogen bonds along the chain. However, the O-H stretching region of the vibrational spectra of cellulose I shows only a broad band similar to that observed for methyl β-cellobioside.

As was demonstrated in the previous section, certain implications concerning the structures of the compounds being investigated may be obtained from consideration of the observed frequencies in the O-H stretching region. Some general comments may be made concerning the structures of the cellodextrins, cellulose I and cellulose II based on the interpretation of the O-H stretching region. Firstly, it may be stated that once a DP of four is reached, there are no significant changes in the O-H stretching region of the vibrational spectra of the cellodextrins and therefore no significant changes in structure. This is consistent with the earlier observation concerning the vibrational spectra below 1500 cm^{-1} and with observations made by Williams[62] concerning the x-ray powder

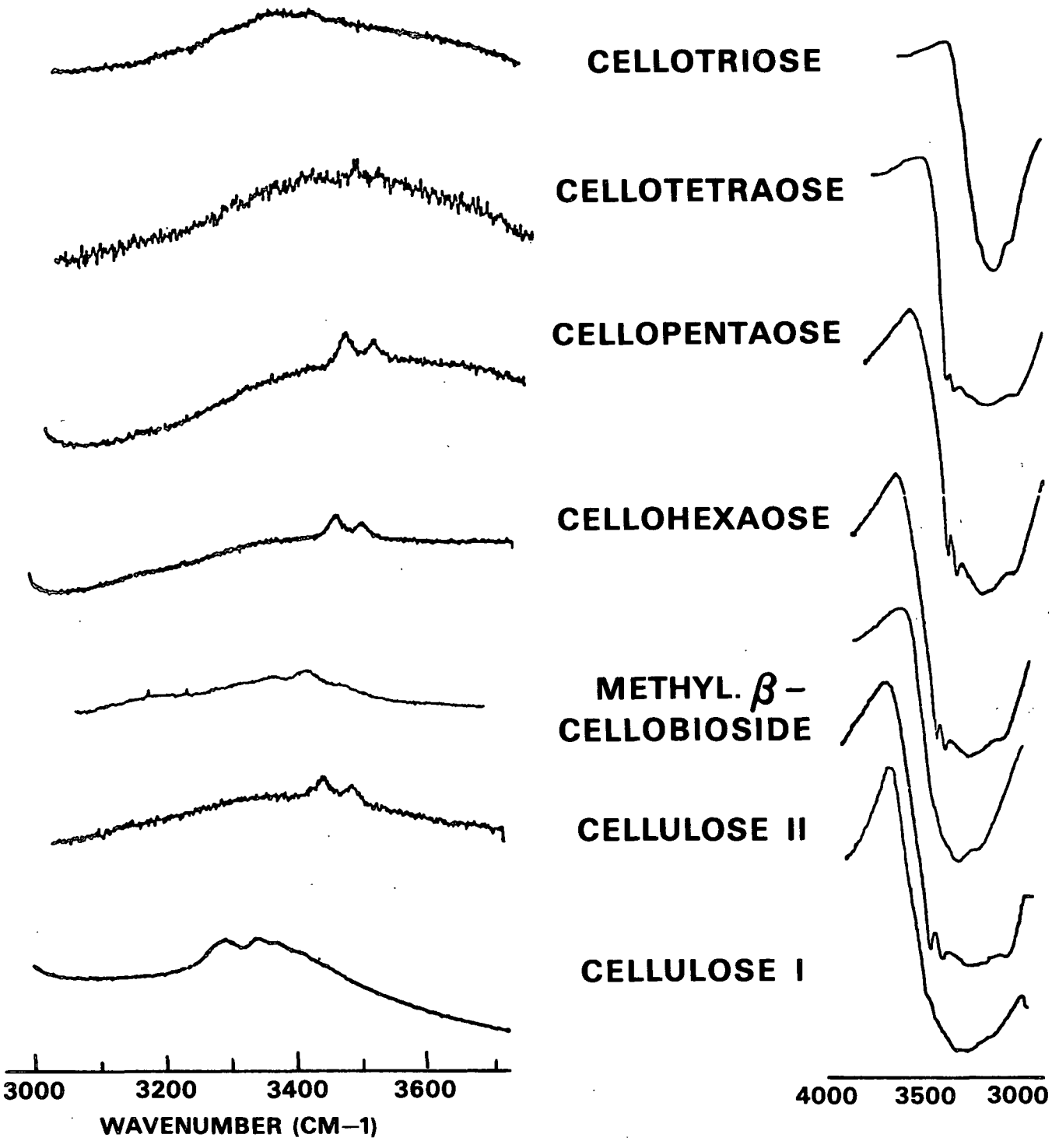


Figure 73. O-H stretching regions of the cellodextrins, methyl β-cellobioside, cellulose I and cellulose II.

patterns of the cellodextrins. Secondly, the existence of two sharp bands in the O-H stretching region of the vibrational spectra of cellulose II and the higher cellodextrins suggests the presence of two nonequivalent intramolecular hydrogen bonds. This suggests that there are probably two nonequivalent conformations about the glycosidic linkages present in these compounds. The O-H stretching region for cellotriose indicates that any intramolecular hydrogen bonds are not isolated. It is not possible to differentiate between the possibilities of the intramolecular bond being bifurcated or perturbed by the intermolecular hydrogen bonding network. Both interpretations are consistent with the observed spectra of cellotriose in the O-H stretching region. The O-H stretching region for cellulose I is very similar to that of methyl β-cellobioside. This suggests that cellulose I contains a bifurcated intramolecular hydrogen bond similar to that found in methyl β-cellobioside.

All of the above observations may be used to construct a coherent description of the structures of the cellodextrins, cellulose I and cellulose II. If the repeating unit for cellulose is taken as cellobiose rather than glucose, either an isolated intramolecular or bifurcated intramolecular hydrogen bond may be included in the repeating unit. This would depend on whether the repeating unit had the conformation of cellobiose or methyl β-cellobioside. A second consideration is the manner in which these repeating units are linked together. The interpretation based on the O-H stretching regions concludes that the cellodextrins of DP greater than four and cellulose II have repeating units of conformations similar to cellobiose and that these repeating units are linked together with glycosidic linkages of conformation also similar to that of cellobiose. This linkage, joining the cellobiose repeating units, would be of slightly different conformation to account for the second sharp observed band. The cellulose I structure would have repeating units with linkage conformations

similar to methyl β-cellobioside which are joined by similar linkages. Thus all the glycosidic linkages in cellulose I would allow the bifurcated intramolecular hydrogen bonds consistent with the observed broad band in the cellulose I O-H stretching region. These results are consistent with the observations based on the normal coordinate analysis of cellobiose and cellotetraose presented earlier.

The O-H stretching region of cellotriose may provide further insight into the nature of the structures of the cellodextrins. The cellotriose O-H stretching region suggests that there is no isolated intramolecular hydrogen bond present. It is possible to view the cellotriose molecule as a derivative of the repeating unit of cellulose, namely cellobiose, and therefore it would not necessarily have to behave as cellobiose and the higher cellodextrins. This could then be applied to all the odd numbered cellodextrins. Viewing the odd numbered DP cellodextrins as derivatives may account for the greater difficulty in getting the odd numbered cellodextrins to crystallize than the even numbered DP cellodextrins.[48]

The O-H stretching region of the cellodextrins, cellulose I and cellulose II may be interpreted in terms of intermolecular and intramolecular hydrogen bonds. Certain implications may be drawn concerning their molecular configurations from these observations. These implications are consistent with the conclusions from the normal coordinate analyses of cellobiose and cellotetraose. These conclusions are that cellulose I has conformations at the glycosidic linkages similar to that found in methyl β-cellobioside and that cellulose II has conformations about the glycosidic linkages similar to that found in cellobiose. The observed differences between the vibrational spectra of cellulose I and cellulose II can therefore be interpreted in terms of conformational differences about the glycosidic linkages.

ACKNOWLEDGMENTS

An endeavor the magnitude of a doctoral dissertation is very seldom a singular effort. There are many people to whom I would like to extend my gratitude.

Firstly, Dr. Rajai Atalla has been much more than a guide or advisor; he has been a friend. The other members of my thesis committee, Dr. G. A. Baum, Dr. L. J. Pitzner, and Dr. R. D. McKelvey, have added many insights into this work.

I would also like to thank The Institute of Paper Chemistry and the member companies for the generous financial assistance during my graduate studies.

A great deal of encouragement during my thesis work has come from my parents, Mr. and Mrs. Enoch J. Carlson, and my parents-in-law, Mr. and Mrs. Allan P. Arthur. Special thanks are warmly extended to my wife, Lynne, who has truly made this work possible through support and self-sacrifice, both financial and emotional. Lastly, I would like to thank the LORD for it is by His will that all things are possible.

LITERATURE CITED

1. Kuhn, L. D., Anal. Chem. 22(2):277-83(1950).

2. Barker, S. A., Bourne, E. J., Stacey, M., and Whiffen, D. H., J.C.S. 1954: 171-6.

3. Barker, S. A., Bourne, E. J., Stacey, M., and Whiffen, D. H., J.C.S. 1954: 3468-73.

4. Barker, S. A., Bourne, E. J., Stacey, M., and Whiffen, D. H., J.C.S. 1954: 4211-15.

5. Neely, W. B., Adv. Carbohyd. Chem. 12:13-33(1957).

6. Spedding, H., Adv. Carbohyd. Chem. 19:23-49(1964).

7. Tipson, R. S., NBS Monograph 110, June, 1968. 21 p.

8. Snyder, R. G. and Schachschneider, J. H., Spectrochim. Acta 19:85-116(1963).

9. Schachschneider, J. H. and Snyder, R. G., Spectrochim. Acta 19:117-68(1963).

10. Snyder, R. G. and Schachschneider, J. H., Spectrochim. Acta 21:169-95(1965).

11. Snyder, R. G. and Zerbi, G., Spectrochim. Acta 23A:391-437(1967).

12. Pickett, L. J. and Strauss, H., J. Chem. Phys. 53:376-88(1970).

13. Vasko, P. D., Blackwell, J., and Koenig, J. L., Carbohyd. Res. 23:407-16 (1972).

14. Cael, J. J., Koenig, J. L., and Blackwell, J., Carbohyd. Res. 32:79-91 (1974).

15. Cael, J. J. Vibrational analysis of polysaccharides. Doctoral Dissertation, Case Western Reserve, Cleveland, Ohio, June, 1976.

16. Cael, J. J., Gardner, K. H., Koenig, J. L., and Blackwell, J., J. Chem. Phys. 62:1145-53(1975).

17. Zhbankov, P. and Sivchik, B. B., Zh. Prikl. Spektr. 26(5):853-9(1977).

18. Chu, S. S. C. and Jeffrey, G. A., Acta Cryst. B24:830-9(1968).

19. Pitzner, L. J. An investigation of the vibrational spectra of the 1,5-anhydropentitols. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, June, 1973.

20. Pitzner, L. J. and Atalla, R. H., Spectochim. Acta 31A:911-29(1975).

21. Fletcher, R. and Powell, M. J. D., Computer J. 6:163-8(1963).

22. Watson, G. M. An investigation of the vibrational spectra of the pentitols and erythritol. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, Jan., 1976.

23. Edwards, S. L. An investigation of the vibrational spectra of the pentose sugars. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, Jan., 1976.

24. Wells, H. A. An investigation of the vibrational spectra of glucose, galactose, and mannose. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, June, 1976.

25. Williams, R. M. An investigation of the vibrational spectra of the inositols. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, June, 1977.

26. Courtesy of Dr. D. G. Williams, 1975.

27. Courtesty of J. Kidd, The Institute of Paper Chemistry.

28. Wilson, E. B., Decius, J. C., and Cross, P. C., Molecular Vibration, New York, McGraw-Hill Book Co., 1955. 388 p.

29. Schonland, D. S., Molecular Symmetry, London, Van Nostrand Reinhold Co., 1965. 298 p.

30. Steele, D., Theory of Vibrational Spectroscopy of Solids, London, Cambridge University Press, 1972. 108 p.

31. Ham, J. T. and Williams, D. G., Acta Cryst. B26(9):1373-83(1970).

32. Fries, D. C., Rao, S. T., and Sundaralingam, M., Acta Cryst. B27:994-1004 (1971).

33. Shirley, S. S. C. and Jeffrey, G. A., Acta Cryst. 23:1038-49(1967).

34. Wells, H. A., unpublished work, 1976.

35. Becker, R. and Sawodny, W., J. Mol. Structure 29:105-15(1975).

36. Cosee, M. and Schachschneider, J. H., J. Chem. Phys. 44:97(1966).

37. Brooks, W. V. F. and Haas, C. M., J. Phys. Chem. 77(3):650-5(1967).

38. Mikawa, Y., Brasch, J. W., and Jakobsen, R., J. Mol. Spectry. 24:314-29 (1967).

39. Fuhrer, H., Kartha, V. B., Krueger, P. J., Mantsah, H. H., and Jones, R. N., Chem. Rev. 72(5):439-56(1972).

40. Michell, A. J., Austral. J. Chem. 28:335-41(1975).

41. Sarko, A. and Muggli, R., Tappi 61(2):45-7(1978).

42. Chanzy, H., Dube, M., and Marchessault, R. H., Tappi 61(7):81-5(1978).

43. Sarko, A. and Muggli, R., Macromol. 7:486-94(1974).

44. Sarko, A., Appl. Polymer Symp. (28):729-42(1976).

45. Zhbankov, R. G., Ivanova, N. V., and Komar, V. P., Vysokomol. Soed. 8(10): 1778-82(1966).

46. Stepanov, B. I., Zhbankov, R. G., and Marupov, R., Vysokomol. Soed. 3:1633-45(1961).

47. Atalla, R. H., Appl. Polymer Symp. (28):659-69(1976).

48. Hayashi, J., Sueko, A., and Watanabe, S., J. Chem. Soc. Japan 7:1320-8 (1977).

49. Atalla, R. H. and Dimick, B. E., Carbohyd. Res. 39:C1-3(1975).

50. Atalla, R. H., Dimick, B. E., and Nagel, S. C., ACS Symp. Ser. (48):30-41 (1977).

51. Dimick, B. E. The importance of the structure of alkali metal hydroxide solutions in decrystallizing cellulose I. Doctoral Dissertation, The Institute of Paper Chemistry, Appleton, Wisconsin, Jan., 1976.

52. Reese, D. A. and Skerrett, R. J., Carbohyd. Res. 7:334(1968).

53. Miller, G. A., In Methods in Carbohydrate Chemistry III - Cellulose, R. H. Whistler, editor, Academic Press, New York and London, 1963. p. 134.

54. Ihat, M. and Goring, D. A., Can. J. Chem. 45(20):2353-8(1967).

55. Ladisch, M. R., Huebner, A. L., and Tsao, G., J. Chromat. 147:185-92 (1978).

56. Gum, E. K. and Brown, R. D., Anal. Biochem. 82:372-5(1977).

57. Dickey, E. E., private communication, 1976.

58. Chanzy, H., private communication, 1977.

59. Kidd, J., private communication, 1977.

60. MacLeod, J. M., private communication, 1977.

61. Hadzi, D. and Bratos, S., In The Hydrogen Bond, P. Shuster, G. Zundel, and C. Sandorfy, editors, North-Holland Publishing Company, Amsterdam, New York, Oxford, 1976. p. 565-612.

62. Williams, D. G., J. Polymer Sci. Part A2 8:637-42(1970).

APPENDIX I

DESCRIPTION OF THE NORMAL COORDINATE CALCULATIONS

INTRODUCTION

The purpose of this Appendix is to describe in detail the mathematics and computer programs which are used in the normal coordinate analysis of the cellodextrins. The order of program descriptions generally follows the order in which the programs would be normally run. The flow of data for the programs is presented in Fig. 74. A master tape of all these programs and some sample input files have been left at the IPC Computer Center.

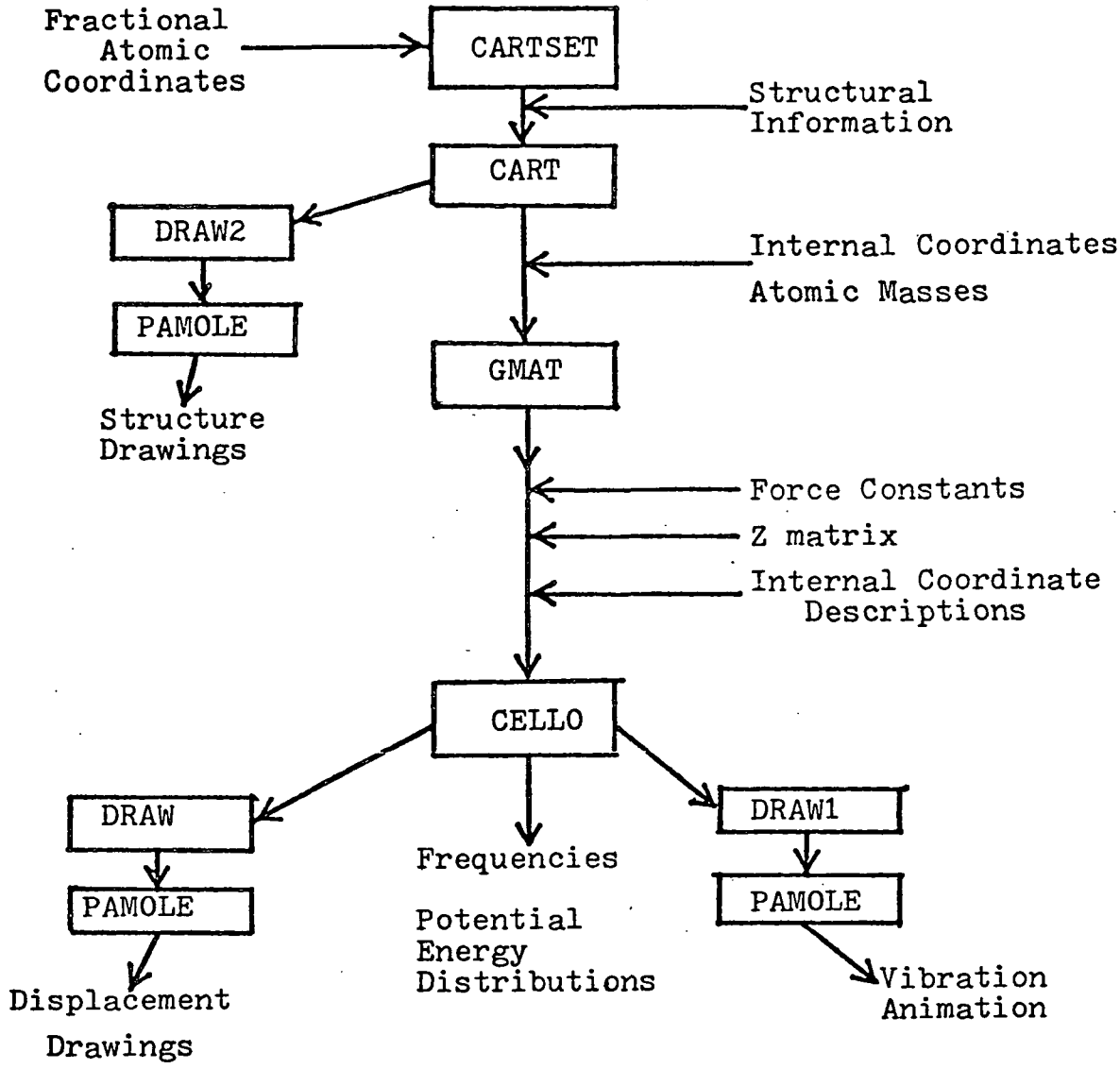


Figure 74. General flow of computer programs.

CARTSET

CARTSET generates the cartesian coordinates and appropriate input data for the program CART from fractional atomic coordinates. The program was written by H. Wells borrowing heavily from NRC-22, a part of IBM's crystallographic library.

First, a transformation matrix is constructed which will transform fractional atomic coordinates into cartesian coordinates. The transformation matrix is:

$$\overline{T} \equiv \begin{pmatrix} A & B \times \cos\zeta & C \times \cos\beta \\ 0 & B \times \sin\zeta & C \times \left(\cos\zeta - \frac{\cos\beta \times \cos\zeta}{\sin\zeta}\right) \\ 0 & 0 & C \times (P) \end{pmatrix}$$

where

A,B,C = lengths of the sides of the unit cell

α, β, ζ = angles between the faces of the unit cell

$$P = \left(\frac{1 + 2 \times \cos\alpha \times \cos\beta \times \cos\zeta - \cos\alpha^2 \times \cos\beta^2 \cos\zeta^2}{\sin\zeta}\right)$$

The fractional atomic coordinates are transformed to cartesian coordinates by:

$$\overline{x}_{cart} = \overline{T} \cdot \overline{x}_{fract}$$

CART input is generated first for atoms 2 and 3 and then for the remaining atoms sequentially. The CART input for each atom consists of 3 reference atoms NA, NB, AND NC, the distance between the atom and NA, the angle between the atom NA and NB, and the angle between the bond of the atom and NA and the plane defined by NA, NB, and NC. The distance NO-NA is taken from the distance table generated previously. The angle NO-NA-NB is calculated by:

$$\beta = (180 - \cos^{-1} (\frac{(NO,NB)^2 - (NO,NA)^2 - (NA,NB)^2}{2x\ (NO,NA)\ x\ (NO,NB)} \times 57.25579)$$

where

(NO,NA) = distance between NO and NA.

A set of direction cosines must be determined before the dihedral angle, θ, between NO-NA and NA-NB-NC can be calculated. The angle θ is then calculated by:

$$\theta = [\cos^{-1} (\cos\alpha_1 \times \cos\alpha_2 + \cos\beta_1 \times \cos\beta_2 + \cos\ \zeta_1 \times \cos\ \zeta_2)]$$

where

α_1, β_1, ζ_1 = direction cosines for NO-NA-NB .

α_2, β_2, ζ_2 = direction cosines for NA-NB-NC

This value and $180-\theta$ are printed. It is up to the operator to determine which value is appropriate and whether it should be positive or negative.

The final portion of the program prints the table of atomic distances.

CART

The program CART calculates cartesian coordinates and atomic distances from bond length, bond angles, and dihedral angles. The program can also calculate the center of mass, the moment of inertia tensor and the principle cartesian coordinates of the atoms, if desired. The cartesian coordinates are calculated in a two step process: (a) transformation of the spherical coordinates defined by the bond lengths and angles to an arbitrary set of cartesian coordinates and (b) transformation of these cartesian coordinates to cartesian coordinates determined by the first three atoms defined.

The fixed cartesian coordinate system is determined by the first 3 atoms in the following manner. Atom 1 determines the origin. The direction from atom 1 to atom 2 determines the positive x axis. Atoms 1, 2, and 3 determine the x,y plane with the positive y direction determined by the position of atom 3 relative to the x axis. The positive z direction is set by requiring a right-handed coordinate system.

In order to determine the cartesian coordinates of atom NO, three atoms, NA, NB, and NC, must be specified for which cartesian coordinates have been previously calculated. Also required are, R, the bond length NO-NA, θ, the angle NO-NA-NB and Φ, the angle between NO-NA and the plane NA-NB-NC. These spherical coordinates are transformed to an arbitrary cartesian coordinate system by:

$$x' = R \cos \theta$$

$$y' = R \sin \theta * \cos \Phi$$

$$z' = R \sin \theta * \sin \Phi$$

These coordinates are then transformed to the set cartesian coordinate system by:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \begin{pmatrix} i \cdot i' & i \cdot j' & i \cdot k' \\ j \cdot i' & j \cdot j' & j \cdot k' \\ k \cdot i' & k \cdot j' & k \cdot k' \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} x_o \\ y_o \\ z_o \end{pmatrix}$$

where

x,y,z = set cartesian coordinates

i,j,k = unit vectors of set cartesian coordinate system

x',y',z' = arbitrary cartesian coordinate system

i',j',k' = unit vectors of arbitrary cartesian coordinate system

x_o, y_o, z_o = cartesian coordinates of atom NA in set cartesian coordinate system.

The values of the dot products of the unit vectors are calculated from the cartesian coordinates of the three specified atoms. First the differences in the cartesian coordinates are determined as:

$$\Delta x_b = x_b - x_a \qquad \Delta x_c = x_c = x_a$$

$$\Delta y_b = y_b - y_a \qquad \Delta y_c = y_c - y_a$$

$$\Delta z_b = z_b - z_a \qquad \Delta z_c = z_c - z_a$$

The first 3 columns of the transformation matrix are calculated in the following manner:

$$i \cdot i' = \Delta x_b/r = \cos\alpha$$

$$j \cdot i' = \Delta y_b/r = \cos\beta$$

$$k \cdot i' = \Delta z_b/r = \cos\zeta$$

$$i \cdot j' = \Delta x_i'/r' = \cos\alpha'$$

$$j \cdot j' = \Delta y_i'/r' = \cos\beta'$$

$$k \cdot j' = \Delta z_i'/r' = \cos\zeta'$$

where

$$r = (\Delta x_b^2 + \Delta y_b^2 + \Delta z_b^2)^{1/2}$$

$$\Delta x_i' = \Delta x_c - (\Delta x_c^* \cos\alpha + \Delta y_c^* \cos\beta + \Delta z_c^* \cos\zeta) \cos\alpha$$

$$\Delta y_i' = \Delta y_c - (\Delta x_c^* \cos\alpha + \Delta y_c^* \cos\beta + \Delta z_c^* \cos\zeta) \cos\beta$$

$$\Delta z_i' = \Delta z_c - (\Delta x_c^* \cos\alpha + \Delta y_c^* \cos\beta + \Delta z_c^* \cos\zeta) \cos\zeta$$

$$r' = \equiv (x_i'^2 + y_i'^2 + z_i'^2)^{1/2}$$

The remaining elements are determined using the following vector identity:

$$i \cdot k' = (j \cdot i')*(k \cdot j') - (j \cdot j')*(k \cdot i') = \cos\alpha''$$

$$j \cdot k' = -(i \cdot i')*(k \cdot j') + (i \cdot j')*(k \cdot i') = \cos\beta''$$

$$k \cdot k' = (i \cdot i')*(j \cdot j') - (i \cdot j')*(j \cdot i') = \cos\zeta''$$

The atomic distances are calculated by:

$$R_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2}$$

A table of the atomic distances is printed as a check of the input data. This information may be checked against the atomic distances printed by CARTSET to assure that proper dihedral angles have been used.

The center of mass is calculated by:

$$x_{cm} = \Sigma_i \; x_i \; wt_i / \Sigma_i wt_i$$

$$y_{cm} = \Sigma_i \; y_i \; wt_i / \Sigma_i wt_i$$

$$z_{cm} = \Sigma_i \; z_i \; wt_i / \Sigma'_i \; wt_i$$

where

wt_i is the mass of atom number i.

At the option of the user, the elements of the moment of inertia tensor are calculated. They are obtained according to:

$$I_{xx} = \Sigma_i \; (y_i^2 + Z_i^2) \; wt_i$$

$$I_{yy} = \Sigma_i \; (z_i^2 + x_i^2) \; wt_i$$

$$I_{zz} = \Sigma_i \; (x_i^2 + y_i^2) \; wt_i$$

$$I_{xy} = I_{yx} = - \; \Sigma_i \; x_i \; y_i \; wt_i$$

$$I_{xz} = I_{zx} = - \; \Sigma_i \; x_i \; z_i \; wt_i$$

$$I_{yz} = I_{zy} = - \; \Sigma_i \; y_i \; z_i \; wt_i$$

The moment of inertia tensor is diagonalized by the Jacobi method in subroutine HDIAG. The principle moments of inertia may be printed at the users option. The transformation which diagonalizes the moment of inertia tensor may be applied to the cartesian coordinates for each atom to yield the principle cartesian coordinates of each atom.

DETERMINATION OF THE B AND G MATRICES

The motions of an isolated molecule may be described in terms of infinitesimal changes in bond lengths, bond angles and torsional or dihedral angles. These motions constitute a set of internal coordinates (i.e., coordinates unaffected by rotation or translation of the molecule as a whole). Since the displacements are assumed to be infinitesimal, only first order linear terms need be considered. This results in considerable simplification of the mathematics involved. One of these internal coordinates S_t, may be described by:

$$S_t = \sum_{i=1}^{3N} B_{ti}\varepsilon_i \qquad T = 1,2,3,\ldots\ 3N-6$$

where the coefficients B_{ti} are determined by the geometry of the molecule and ε_i represents a cartesian displacement coordinate. Wilson Decius and Cross introduce the following notation

$$S_t = \sum_{\alpha=1}^{N} \overline{s}_t\ \alpha \cdot \overline{\rho}a$$

There is a vector, $\rho\alpha$, for each atom α whose components are $\varepsilon_{3\alpha}$, $\varepsilon_{,3\alpha+1}$, $\varepsilon_{3\alpha+2}$. The vector $s_{t\alpha}$ may be visualized as lying in the direction which will produce the largest increase in $\overline{S}_t$ when only atom α is moved (all other atoms are in their equilibrium positions). It has the magnitude, $s_{t\alpha}$, equal to the increase in S_t produced by a unit displacement in this most effective direction.

CONSTRUCTION OF THE B MATRIX

In the computer program GMAT the components of the B matrix are calculated in one of three subroutines depending on the type of internal coordinate being considered. The mathematics are described in this section. The unit vector

pointing from atom 1 to atom 2 will be denoted as $\bar{e}_{12}$ in the following descriptions.

BOND STRETCHING

In bond stretching, S_t is the increase in the distance between atoms 1 and 2 as shown in Fig. 75. From the definition of $\bar{s}_t$, it is clear that $\bar{s}_{t1}$ and $\bar{s}_{t2}$ lie along the line connecting atoms 1 and 2 directed away from the opposing atom. The elements of the vector $\bar{s}_{t\alpha}$ for a bond stretching internal coordinate for atoms 1 and 2 are:

$$\bar{s}_{t1} = \bar{e}_{21} = -\bar{e}_{12} \quad \text{and} \quad \bar{s}_{t2} = \bar{e}_{12} = -\bar{e}_{21}$$

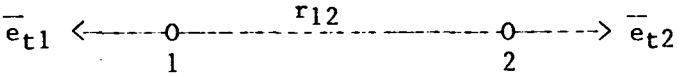


Figure 75. Diagram of bond stretching internal coordinate.

The calculation of the $\underline{B}$ matrix elements associated with the bond stretching internal coordinates is performed in the subroutine BOST.

VALENCE ANGLE BENDING

Valence angle bending is described as the increase in the angle between two valence bonds attached to the same atom 3 as shown in Fig. 76. Consider $\bar{s}_{t1}$ first. The greatest increase in the angle will result from a displacement of atom 1 perpendicular to bond 3-1 away from atom 2. The increase in the angle for a unit displacement in this direction is $1/r_{31}$ where r_{31} is the distance from 3 to 1. This is true because we are considering only infinitesimal displacements. By a similar argument $\bar{s}_{t2}$ will be pointed away from 1 perpendicular to bond 3-2. It will have a magnitude of $1/r_{23}$.

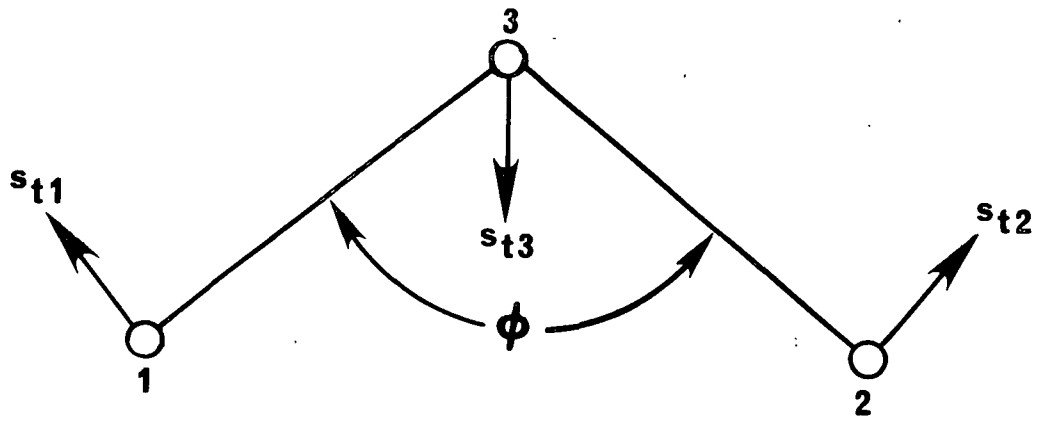


Figure 76. Diagram of valence angle bending internal coordinate.

It can be shown that the sum of all the s vectors associated with a single internal coordinate must equal zero. Thus $\bar{s}_{t3}$ must be the negative of the sum of $\bar{s}_{t1}$ and $\bar{s}_{t2}$. These s vectors may be expressed in terms of the unit vectors $\bar{e}_{31}$ and $\bar{e}_{32}$ in the following manner.

$$\bar{s}_{t1} = \frac{\cos\theta\,\bar{e}_{31} - \bar{e}_{32}}{r_{31}\sin\theta}$$

$$\bar{s}_{t2} = \frac{\cos\theta\,\bar{e}_{32} - \bar{e}_{31}}{r_{32}\sin\theta}$$

$$\bar{s}_{t3} = \frac{(r_{31} - r_{32}\cos\theta)\,\bar{e}_{31} + (r_{32} - r_{31}\cos\theta)\,\bar{e}_{32}}{r_{31}\,r_{32}\sin\theta}$$

An alternate method using vector products is presented by Wilson Decius and Cross for obtaining the s vectors associated with the valence angle bending internal coordinate. The $\underline{B}$ matrix elements associated with the valence angle bending internal coordinates are calculated in the subroutine BEND.

TORSIONS

A torsional internal coordinate is defined as the change in the dihedral angle, τ, between the planes defined by atoms 1, 2, and 3 and atoms 2, 3, and 4

when atoms 1, 2, 3, and 4 are bonded as shown in Fig. 77. A sign convention is adopted for the angle τ. The absolute value of τ is always less than π and is said to be positive if atom 4 lies clockwise from atom 1 when viewed down the bond connecting atoms 2 and 3 from 2 to 3.

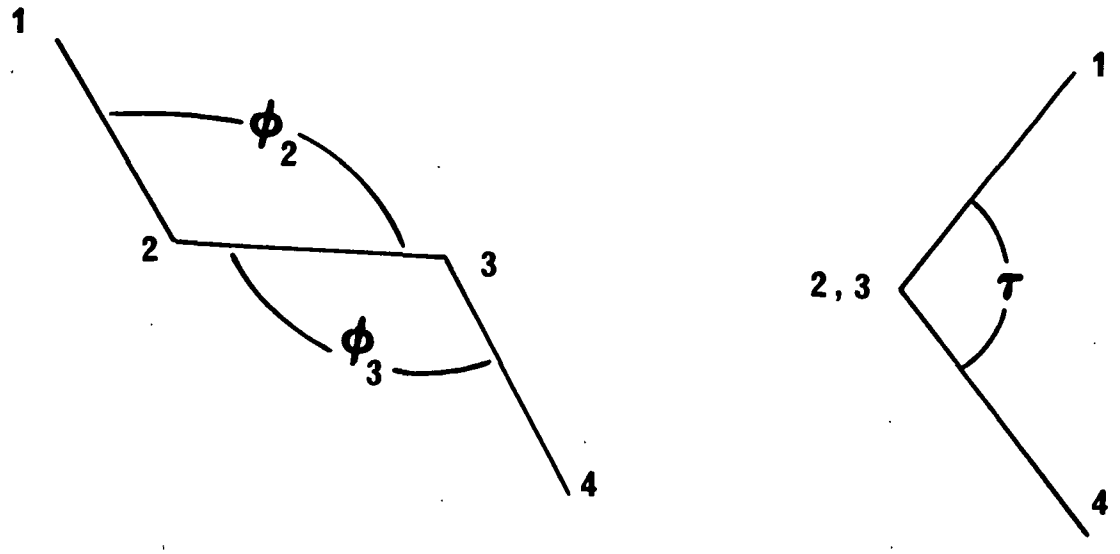


Figure 77. Diagram of torsional internal coordinate.

The unit vectors perpendicular to the planes defined by 1, 2, and 3 and 2, 3, and for are, respectively, $(\bar{e}_{34} \times \bar{e}_{32})/\sin\theta_3$. Using this, the value of τ can be expressed as:

$$\tau = \cos^{-1} \frac{(\bar{e}_{12} \times \bar{e}_{23}) \cdot (\bar{e}_{23} \times \bar{e}_{34})}{\sin\theta_2 \sin\theta_3}$$

First consider the elements s_{t1}, the contribution of atom 1 to S_t. The most effective direction for an <u>increase</u> in τ is $(\bar{e}_{21} \times \bar{e}_{23})$ and a unit vector in this direction would be $(\bar{e}_{21} \times \bar{e}_{23})/\sin\theta_2$. An <u>infinitesimal</u> displacement in this direction would increase by $(r_{12} \sin\theta_2)^{-1}$. Thus $\bar{s}_{t1}$ takes the form:

$$\bar{s}_{t1} = \frac{1}{r_{12} \sin\theta_2} \frac{(\bar{e}_{21} \times \bar{e}_{23})}{\sin\theta_2} = -\frac{(\bar{e}_{12} \times \bar{e}_{23})}{r_{12} \sin^2\theta_2}$$

Similarly

$$\bar{s}_{t4} = - \frac{(\bar{e}_{34} \times \bar{e}_{32})}{r_{34} \sin^2 \theta_2}$$

A movement of atom 2 results in two components which change the value of τ. One component is perpendicular to plane 1, 2, 3 and the other is perpendicular to plane 2, 3, 4. The component perpendicular to plane 1, 2, 3 has a unit vector of

$$(\bar{e}_{21} \times \bar{e}_{23})/\sin \theta_2$$

and a value of

$$- \frac{1}{r_{12} \sin \theta_2} + \frac{\bar{e}_{12} \cdot \bar{e}_{23}}{r_{23} \sin \theta_2} = - \frac{r_{23} - r_{12} \cos \theta_2}{r_{23} r_{12} \sin \theta_2}$$

The component of $\bar{s}_{t2}$ perpendicular to the plane 2, 3, 4 has the unit vector

$$(\bar{e}_{32} \times \bar{e}_{34})/\sin \theta_3$$

and a value of

$$- \frac{\bar{e}_{23} \cdot \bar{e}_{34}}{r_{23} \sin \theta_3} = - \frac{\cos \theta_3}{r_{23} \sin \theta_3}$$

Therefore, the s vector for atom 2 may be expressed as

$$\bar{s}_{t2} = - \frac{(r_{23} - r_{12} \cos \theta_2) (\bar{e}_{21} \times \bar{e}_{23})}{r_{23} r_{12} \sin^2 \theta_2} - \frac{\cos \theta_3 (\bar{e}_{32} \times \bar{e}_{34})}{r_{23} \sin^2 \theta_3}$$

Similarly

$$\bar{s}_{t3} = - \frac{(r_{32} - r_{43} \cos \theta_3) (\bar{e}_{34} \times \bar{e}_{32})}{r_{32} r_{43} \sin^2 \theta_3} - \frac{\cos \theta_2 (\bar{e}_{21} \times \bar{e}_{23})}{r_{32} \sin^2 \theta_2}$$

The B matrix elements associated with the torsional internal coordinates are calculated in the subroutine TORS.

CONSTRUCTION OF THE G MATRIX

The element, $G_{tt'}$, of the $\bar{G}$ matrix may be defined by

$$G_{tt'} = \sum_{i=1}^{3N} \frac{1}{m_i} B_{ti} B_{t'i}$$

where m_i is the mass of atom i. The elements of the G matrix may also be expressed in terms of the $\bar{s}$ vectors described in the previous sections as

$$G_{tt'} = \sum_{i=1}^{N} \frac{1}{m_i} \bar{s}_{ti} \cdot \bar{s}_{t'i}$$

The program GMAT utilizes upper equation to determine the elements of the G matrix. Clearly, the G matrix is a symmetric matrix. The program takes advantage of this in that it calculates only values for the upper triangle of the G matrix. In the form used in this thesis work, the program GMAT wrote the B and G matrices on a tape which was later used to set up the input data for the solution of the secular equation.

CELLO

The program CELLO calculates the theoretical frequencies, potential energy distributions and the cartesian displacement coordinates for a molecule. The input data for this program are the force constants, the Z matrix, the G matrix, cartesian coordinates and the B matrix for the molecule being considered.

POTENTIAL AND KINETIC ENERGY OF THE MODEL

It is necessary to first develop the theory for the solution of the equations of motion for an isolated vibrating molecule. The potential energy of the model, V, may be expressed in terms of a Taylor's series in the 3N-6 internal coordinates as in Eq. (1).

$$2V = 2V_o + 2 \sum_{i=1}^{3N} (\partial V/\partial q_i)_o \, q_i + \sum_{i,j=1}^{3N} (\partial^2 V/\partial q_i \, \partial q_j) q_i q_j + \text{higher terms} \tag{1}$$

$$= 2V_o + 2 \sum_{i=1}^{3N} f_i \, q_i + \sum_{i,j=1}^{3N} f_{ij} \, q_i q_j + \text{higher terms}$$

The potential energy at equilibrium may be set equal to zero, resulting in $V_o = 0$. The terms f_i are the first derivatives of the potential energy function at the equilibrium position. Since the potential energy function is a minimum at the equilibrium position, these terms, f_i, are zero. For small displacements, the higher order terms may be neglected. This allows the potential energy of the model to be written as in Eq. (2) where $f_{ij} = (\partial^2 V_o/\partial q_i \, \partial q_j)$.

$$2V = \sum_{i,j=1}^{3N} f_{ij} \, q_i q_j \tag{2}$$

The values of the f_{ij}'s are constants which are referred to as force constants. The potential energy is expressed in matrix notation in Eq. (2a) where $\underline{q}$ is a column matrix of the q_i's and $\underline{f}$ is a symmetric matrix with the elements f_{ij}.

$$2V = \underline{q}' \, \underline{f} \, \underline{q} \tag{2a}$$

The kinetic energy of the model may be expressed as

$$2T = \sum_{i,j=1}^{3N} t_{ij}\, \dot{q}_i\, \dot{q}_j \tag{3}$$

where the t_{ij}'s are the kinetic energy coefficients and the $\dot{q}_i$'s are the velocities in terms of the internal coordinates. In matrix notation, this may be expressed as

$$2T = \underline{\dot{q}}\ \underline{t}\ \underline{\dot{q}} \tag{3a}$$

where $\underline{\dot{q}}$ is a column matrix of the velocities $\dot{q}_i$ and $\underline{t}$ is a matrix of the kinetic energy coefficient t_{ij}. In the approximation used in these calculations, the potential and kinetic energies are homogeneous quadratic functions of the generalized coordinates and generalized velocities, respectively.

SETTING UP THE SECULAR EQUATION

In order to solve the vibrational problem for the model which has been chosen, the equations of motion must be developed. This is done by using the Lagranian function and Lagrange's equation of motion. The Lagrangian function, L, for the model may be written as

$$L = T - V = 1/2 \sum_{i,j=1}^{3N} (t_{ij}q_iq_j + f_{ij}q_iq_j) \tag{4}$$

Lagrange's equations of motion are defined by

$$\frac{d\ (\partial L/\partial \dot{q}_j)}{dt} - (\partial L/\partial q_j) = 0 \tag{5}$$

where j = 1,2,3...3N-6. Substitution of Eq. (4) into Eq. (5) yield the equations of motion for the model as

$$\sum_{i=1}^{3N} (t_{ij}\ddot{q}_i + f_{IJ}q_i) = 0 \tag{6}$$

where $j = 1,2,3,\ldots 3N-6$. Equation (6) is a form of the secular equation for the vibrations of an isolated molecule.

It is easier to express and solve the secular equation when it is expressed in matrix form rather than a series of equations as in Eq. (6). The matrix expression for Eq. (6) is

$$\underline{G}^{-1}\,\underline{\ddot{q}} + \underline{F}\,\underline{q} = 0 \tag{7}$$

where $\underline{\ddot{q}}$ and $\underline{q}$ are column matrices of $\ddot{q}_i$ and q_i, $\underline{G}^{-1}$ is the kinetic energy matrix and $\underline{F}$ is the matrix whose elements are f_{ij}. In the calculations, the $\underline{F}$ matrix is constructed from a set of force constants and a matrix which places these force constants in the $\underline{F}$ matrix according to

$$f_{ij} = Z_{ijk}\ \theta\ k \text{ or } \underline{F} = \underline{Z}\ \underline{\theta} \tag{8}$$

where $\underline{\theta}$ is the column matrix of the force constants.

The nontrivial solutions of Eq. (7) must also satisfy

$$\left| F - F^{-1}\ \lambda \right| = 0 \tag{9}$$

or

$$\left| GF - E\ \lambda \right| = 0 \tag{9a}$$

It can be seen that any method which diagonalizes the $\underline{GF}$ product will yield the calculated frequencies. The resulting diagonal elements are referred to as the eigen values of the $\underline{GF}$ product and the columns of the matrix which diagonalizes the $\underline{GF}$ product are called the eigen vectors. This diagonalizing matrix is termed the L matrix and

$$\underline{L}^{-1}\ \underline{G}\ \underline{F}\ \underline{L} = \underline{\Lambda} \tag{10}$$

where $\underline{\Lambda}$ is the diagonal matrix of the eigen values λ_k. The $\underline{L}$ matrix will be shown to have numerous and diverse applications in the normal coordinate calculations.

The $\underline{GF}$ product is generally an asymmetric matrix. This fact complicates the diagonalization. To circumvent this difficulty, a transformation matrix is constructed which allows the transformation of the $\underline{F}$ matrix into a symmetric matrix which has the same eigen values and eigen vectors as the $\underline{GF}$ product.

The transformation matrix which is desired is one which will transform the $\underline{G}$ matrix to a unit matrix. The $\underline{G}$ matrix may be diagonalized in the following manner. First, $G_{i1}G_{1j}/G_{11}$ is subtracted from each element in the G matrix except G_{11} and the resulting matrix is denoted $\underline{G}^{(1)}$. The matrix $\underline{G}^{(1)}$ is symmetric and the first row vanishes. This operation is carried out for rows two through 3N-6, for G_{ij}, $i > j$. The final diagonalized matrix is denoted $\underline{G}^{(m-1)}$ and can be made a unit matrix by

$$\underline{E} = [\underline{G}^{(m-1)}]^{-1/2} \; [\underline{G}^{(m-1)}] \; [\underline{G}^{(m-1)}]^{-1/2} \tag{11}$$

The inverse of the matrix which will diagonalize the G matrix is

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ G_{12}/G_{11} & 1 & & & \cdots \\ G_{13}/G_{11} & G_{23}^{(1)}/G_{22}^{(1)} & 1 & & \cdots \\ G_{14}/G_{11} & G_{24}^{(1)}/G_{22}^{(1)} & G_{34}^{(2)}/G_{33}^{(2)} & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \tag{12}$$

The transformation of $\underline{G}$ is then

$$\underline{G}^{(m-1)} = \underline{A}\,\underline{G}\,\underline{A}' \tag{13}$$

This matrix can then be made a unit matrix as shown in Eq. (11). The corresponding F matrix transformations are

$$\overline{\underline{F}} = (\underline{A}^{-1})'\,\overline{\underline{F}}\,(\underline{A}^{-1}) \tag{14}$$

and

$$\underline{H} = [\underline{G}^{(m-1)}]^{1/2}\overline{\underline{F}}\,[\underline{G}^{(m-1)}]^{1/2} \tag{15}$$

The matrix $\underline{H}$ is a symmetric matrix having eigen values and eigen vectors identical to the $\underline{GF}$ product. Since $\underline{H}$ is symmetric, it can be diagonalized by an orthogonal matrix, $\underline{W}$, such that

$$\underline{H}\,\underline{W} = \underline{W}\,\lambda \text{ and } \underline{W}'\,\underline{H}\,\underline{W} = \underline{E}\,\lambda = \underline{\Lambda} \tag{16}$$

where $\underline{\Lambda}$ is the diagonal matrix of the eigen values, λ_k. Therefore the transformations which diagonalize the kinetic and potential energy matrices are

$$\underline{W}'[\underline{G}^{(m-1)}]^{-1/2}\,\underline{AGA}'\,[\underline{G}^{(m-1)}]^{-1/2}\,\underline{W} = \underline{E} \tag{17}$$

and

$$\underline{W}'[\underline{G}^{(m-1)}]^{1/2}(\underline{A}^{-1})\underline{F}\,\underline{A}^{-1}[\underline{G}^{(m-1)}]^{1/2}\,\underline{W} = \underline{\Lambda} \tag{18}$$

$$\underline{L} = \underline{A}^{-1}[\underline{G}^{(m-1)}]^{1/2}\underline{W} \tag{19}$$

$$\underline{L}^{-1} = \underline{W}'[\underline{G}^{(m-1)}]^{-1/2}\underline{A} \tag{20}$$

From these equations, the computer program CELLO calculates the normal coordinates, theoretical frequencies, potential energy distributions and the cartesian displacement coordinates.

NORMAL COORDINATES

There exists a set of 3N-6 coordinates, Q_i, for each solution of the secular equation such that there are no cross terms between the coordinates. They are called the normal coordinates.

$$Q_i \cdot Q_j = \begin{cases} Q_i^2 & i = j \\ 0 & i \neq j \end{cases} \tag{21}$$

The matrix $\underline{L}$ calculated above provides the transformation from internal to normal coordinates such that

$$q_i = \sum_{j=1}^{3N-6} L_{ij} Q_j \tag{22}$$

or in matrix form where $\underline{Q}$ is a column matrix of the Q_j's

$$\underline{q} = \underline{L}\,\underline{Q} \tag{23}$$

The kinetic energy in terms of normal coordinates may be expressed as in Eq. (24) and the potential energy as in Eq. (25).

$$2T = \dot{\underline{q}}'\underline{T}\,\dot{\underline{q}} = \dot{\underline{Q}}'\underline{L}'\ \underline{t}\ \underline{L}\,\dot{\underline{Q}} = \dot{\underline{Q}}'\ \underline{E}\,\dot{\underline{Q}} = \sum_k \dot{Q}_k \tag{24}$$

$$2V = \underline{q}'\underline{fq} = \underline{Q}'\underline{L}'\ \underline{f}\ \underline{LQ} = \underline{Q}'\ \underline{\Lambda}\,\underline{Q} = \sum_k \lambda_k q_k \tag{25}$$

$\underline{E}$ is the unit matrix and $\underline{\Lambda}$ is the diagonal matrix of the eigen values λ_k.

VIBRATIONAL WAVE EQUATION

It can be demonstrated that the classical mechanical model used in the normal coordinate calculations is consistent with the quantum mechanical interpretation of the vibrational spectra. Vibrational spectra may be interpreted as

the absorption of quanta of energy (h ν_i) associated with the transitions between the ground vibrational state and the excited vibrational states* of wavenumber ν_i. The excited vibrational state can be described by the Schroedinger equation for that state in terms of the internal coordinates.

$$\bar{\psi}_j = \sum_{i=1}^{3N-6} \psi_{ij}(q_i) \tag{26}$$

However, when the transformation is made to normal coordinates each vibration of the molecule is described by a single coordinate and

$$\bar{\psi}_j = \psi_j\ (Q_j) \tag{27}$$

and the entire vibrational wave equation may be written as

$$\bar{\psi}_v = \bar{\psi}_1\ (Q_1)\ \bar{\psi}_2\ (Q_2)\ \ldots\ldots\ \bar{\psi}_{3N-6}\ (Q_{3N-6}) \tag{28}$$

This is a set of 3N-6 one dimensional equations of the form

$$(\partial^2\bar{\psi}_k/\partial Q_k{}^2) + (8\ \pi^2/h^2)\ (E_k = 1.2\ \lambda_k Q_k{}^2) = 0 \tag{29}$$

for k = 1,2,3,...3N-6. Equation (29) can be recognized as the wave equation for a simple harmonic oscillator. The total vibrational energy associated with the wave Eq. (28) is the sum of the E_k's for Eq. (29) associated with the individual normal coordinates.

$$E_{vib} = \sum_{k=1}^{3N-6} E_k \tag{30}$$

*Transitions also occur between the excited states which result in the absorption of quanta of energy other than the amounts associated with the transition from the ground state to individual excited states. However, these alternate transitions do not occur nearly as often and may be neglected as a first approximation.

The solution of Eq. (29) has been well established and the energy levels are given by $(v_k + 1/2)\ h\nu_k$ where v_k is the vibrational quantum number and ν_k is the classical frequency of vibration, $\nu_k = (\lambda_k)^{1/2}/2\pi h$. In this manner the classical mechanical and quantum mechanical interpretations of the phenomenon of vibrational frequencies are consistent and complementary.

CALCULATION OF POTENTIAL ENERGY DISTRIBUTION

The contributions to the potential energy from each diagonal force constant is calculated by the program CELLO. The equation used to calculate these potential energy contributions is shown in Eq. (31).

$$PE_j = L_{ji}{}^2 F_{jj}/\lambda_i \tag{31}$$

The potential energy contribution and the description of the internal coordinate j are printed for all contributions for each calculated vibrations. This information is called the potential energy distribution (PED). The off-diagonal force constants are not considered in this calculation. The extent to which the sum of the PED varies from 100% is an indication of the amount of contribution from the off-diagonal interaction force constants.

CALCULATION OF THE CARTESIAN DISPLACEMENT COORDINATES

The $\underline{L}$ matrix has one more very important application in the normal coordinate calculations, namely the calculation of cartesian displacement coordinates. There exists a transformation $\underline{T}$ from normal coordinates, $\underline{Q}$, to cartesian coordinates $\underline{X}$ such that

$$\underline{X} = \underline{T}\ \underline{Q} \tag{32}$$

This matrix, $\underline{T}$, may be obtained from the relationship

$$\underline{T} = \underline{M}^{-1}\ \underline{B}'\ (\underline{L}^{-1}) \qquad (33)$$

where $\underline{M}^{-1}$ is a matrix of inverse atomic masses and $\underline{B}'$ is the transpose of the $\underline{B}$ matrix calculated by GMAT. The $\underline{T}$ matrix contains the cartesian displacement (Δx, Δy, Δz) for each atom for each normal coordinate. For example, a unit displacement in normal coordinate Q_i will result in the cartesian displacements

$$\underline{X}^i = \underline{T}^i \qquad (34)$$

where $\underline{T}^i$ is the ith column of the T matrix.

There are two options for displaying the cartesian displacement coordinates for a particular mode. The first is to draw a "ball and stick" diagram of the molecule with the displacements represented by "arrows." This is done using the drawing program PAMOLE with the cartesian coordinates of the displacements determined by

$$\underline{X}_d{}^i = \underline{X}_o + \underline{T}^i \qquad (35)$$

where $\underline{X}_o$ is a column matrix composed of the equilibrium cartesian coordinates. The second option is to use PAMOLE to draw a series of pictures of the molecule, each displaced by an increasing fraction of the total displacement for that normal coordinate. This series of pictures may then be animated to demonstrate the type(s) of motions present in a particular normal mode.

PAMOLE

The program PAMOLE draws "ball and stick" representations of the molecule from information concerning the size of each atom its cartesian coordinates and the manner in which the atoms are connected. This program was used for three

purposes in this thesis work. The first was to obtain a diagram of the structures used in the calculations of theoretical frequencies to assure that they were correct. The second was to draw pictures of the displacements for normal modes of particular interest. The final use was to obtain a series of pictures representing the motions associated with a particular normal mode so that the motion might be animated.

UTILITY PROGRAMS

During the course of the thesis work, four utility programs were written. Three of these programs take output from the appropriate source and set up a data file to be used in connection with the program PAMOLE. These three programs are DRAW (for animation), DRAW1 (for cartesian displacements), and DRAW2 (utilizing CART output). The fourth utility program DACOPY, sets up the input data for CELLO from card input of the force constants, internal coordinate descriptions, and Z matrix and tape input of the G and B matrices as written by the program GMAT. These programs are well-documented with comments so that they are essentially self-explanatory.

All of the programs in this Appendix and a set of representative inputs are on a master library tape at the Computer Center of The Institute of Paper Chemistry.

ADDENDUM

Subsequent to the completion of the thesis work, commercial samples of crystalline cellotriose, cellotetraose, cellopentaose, and cellohexaose were obtained from V-Labs, Inc. (Covington, LA). These samples were dissolved in ethanol/water and exposed to acetone vapors to further enhance crystallization in a manner similar to that used in the thesis work. Both the original spectra and the V-Labs spectra are presented here for comparison.

250 cm^{-1} to 1650 cm^{-1}

The cellotriose spectra of the V-Labs cellotriose is much cleaner than that obtained in the thesis work. There are several bands in the original spectrum which are not present in the spectrum for the V-Labs sample. This leads to the conclusion that the original sample was either an anomeric mixture or had trace impurities. The spectrum of the V-Labs sample should be taken as the more representative spectrum for cellotriose. This conclusion does not alter the thesis results as they were not based on cellotriose vibrational analyses.

The cellotetraose and cellopentaose spectra are very similar for both sources with the V-Labs spectra appearing slightly more crystalline.

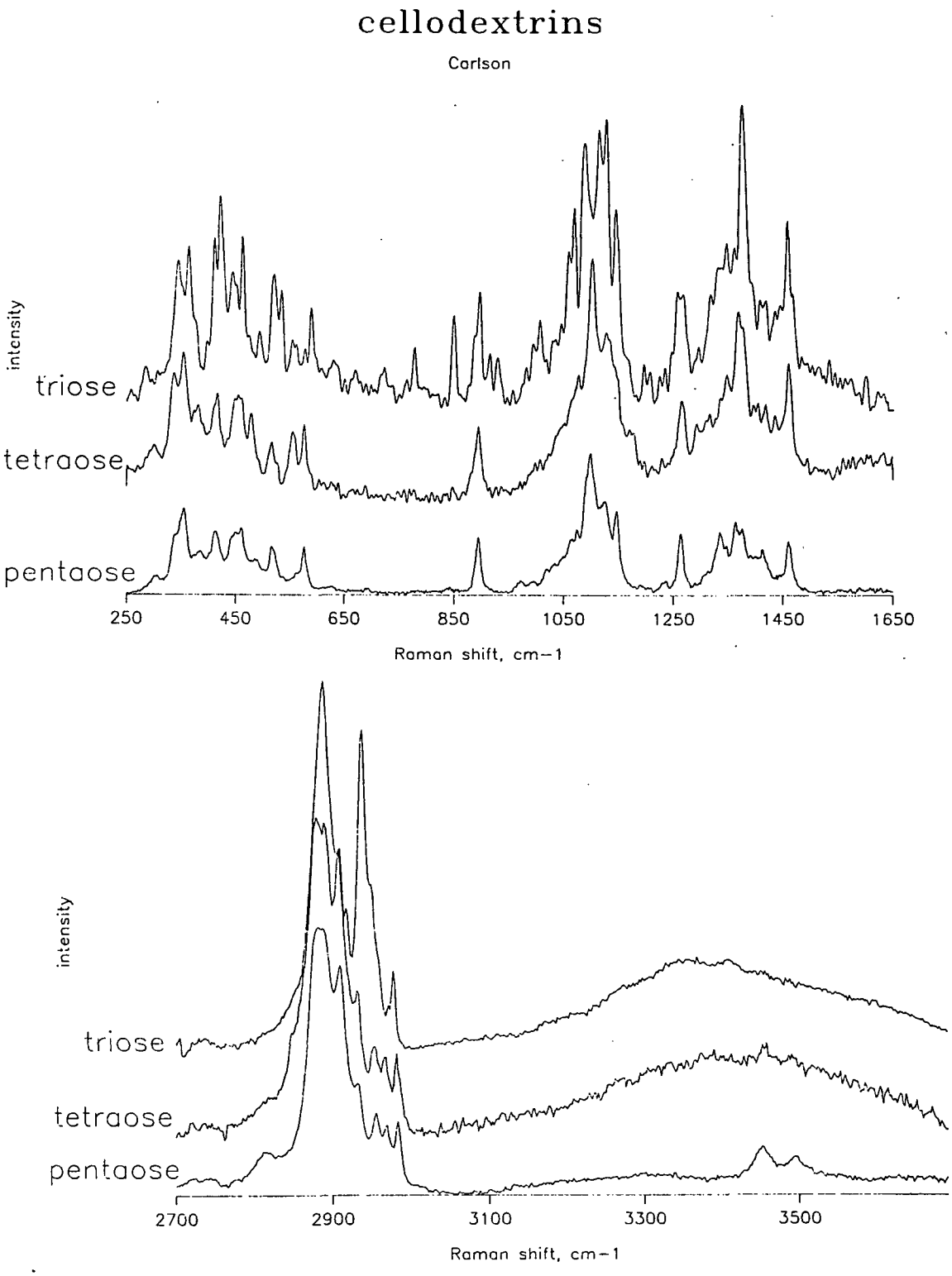
2700 cm^{-1} to 3500 cm^{-1}

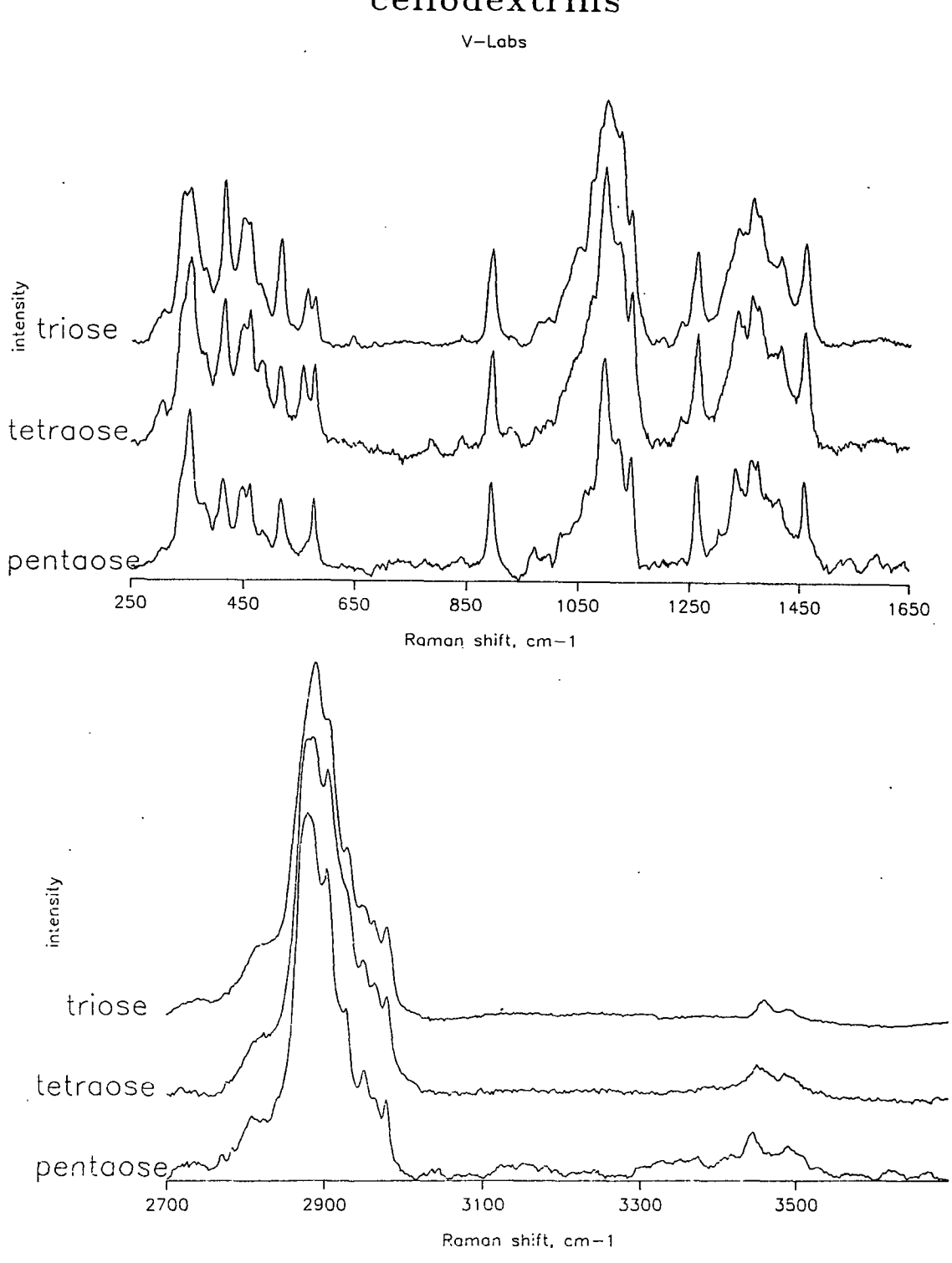
This region was used to support the hypothesis that cellulose I had one type of glycosidic linkage, and cellulose II had two dissimilar alternating linkages. This was based on a diffuse OH stretching region (3100 cm^{-1} to 3500 cm^{-1}) for cellulose I and two sharp bands between 3450 cm^{-1} and 3500 cm^{-1} for cellulose II. Further, the lower DP cellodextrins were believed to have a mixture of linkages and it was not until a DP of 5 (cellopentaose), did the presence of dissimilar linkages become evident from the Raman spectra.

The enhanced purity and crystallinity of the V-Labs samples allow bands which were buried in background "noise" to be evident in both cellotriose and cellopentaose between 3450 cm^{-1} and 3500 cm^{-1}. This extends the concept of dissimilar glycosidic linkages down to cellotriose and further enhances the interpretation of the difference between cellulose I and cellulose II being the conformational differences at the glycosidic linkages.

Conclusions

Improvements in purity and crystallinity of commercial samples of cellotriose, cellotetraose, and cellopentaose have resulted in improved Raman spectra relative to those obtained in the thesis work. These spectra reinforce the thesis conclusion that differences in conformation about the glycosidic linkages can account for the observed differences between cellulose I and cellulose II. Their improved clarity allows the extension of the concept of dissimilar glycosidic linkages to cellotetraose and cellotriose.


cellodextrins
Carlson
intensity
triose
tetraose
pentaose
250
450
650
850
1050
1250
1450
1650
Raman shift, cm−1
intensity
triose
tetraose
pentaose
2700
2900
3100
3300
3500
Raman shift, cm−1


cellodextrins
V-Labs
intensity
triose
tetraose
pentaose
250
450
650
850
1050
1250
1450
1650
Raman shift, cm-1
intensity
triose
tetraose
pentaose
2700
2900
3100
3300
3500
Raman shift, cm-1